

Duality of a Source Coding Problem and the Semi-Deterministic Broadcast Channel with Rate-Limited Cooperation

Ziv Goldfeld, Haim H. Permuter and Gerhard Kramer

Abstract

The Wyner-Ahlsvede-Körner (WAK) empirical-coordination problem where the encoders cooperate via a finite-capacity one-sided link is considered. The coordination-capacity region is derived by combining several source coding techniques, such as Wyner-Ziv (WZ) coding, binning and superposition coding. Furthermore, a semi-deterministic (SD) broadcast channel (BC) with one-sided decoder cooperation is considered. Duality principles relating the two problems are presented, and the capacity region for the SD-BC setting is derived. The direct part follows from an achievable region for a general BC that is tight for the SD scenario. A converse is established by using telescoping identities. The SD-BC is shown to be operationally equivalent to a class of relay-BCs (RBCs) and the correspondence between their capacity regions is established. The capacity region of the SD-BC is transformed into an equivalent region that is shown to be dual to the admissible region of the WAK problem in the sense that the information measures defining the corner points of both regions coincide. Achievability and converse proofs for the equivalent region are provided. For the converse, we use a probabilistic construction of auxiliary random variables that depends on the distribution induced by the codebook. Several examples illustrate the results.

Index Terms

Channel and source duality, Cooperation, Empirical coordination, Multiterminal source coding, Relay-broadcast channel, Semi-deterministic broadcast channel.

I. INTRODUCTION

Cooperation can substantially improve the performance of a network. A common form of cooperation permits information exchange between the transmitting and receiving ends via rate-limited links, generally referred to as

The work of Z. Goldfeld and H. H. Permuter was supported by the Israel Science Foundation (grant no. 684/11) and an ERC starting grant. The work of G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research.

This paper was presented in part at the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, and in part at the 2014 IEEE 28-th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, November, 2014, and was accepted to be presented in part at the 2015 IEEE Information Theory Workshop, Jerusalem, Israel.

Z. Goldfeld and H. H. Permuter are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel (gziv@post.bgu.ac.il, haimp@bgu.ac.il). G. Kramer is with the Institute for Communications Engineering, Technische Universität München, Munich D-80333, Germany (gerhard.kramer@tum.de).

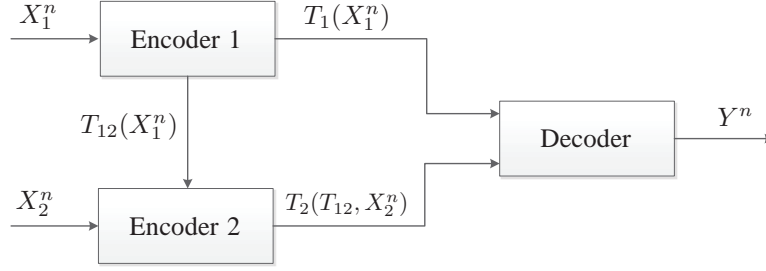


Fig. 1: The WAK source coding problem.

conferencing [1]. In this work, conferencing is incorporated in a special case of the fundamental two-encoder multiterminal source coding problem (cf., e.g., [2] and [3]). Solutions for several special cases of the two-encoder source coding problem have been provided. Among these are the Slepian-Wolf (SW) [4], Wyner-Ziv (WZ) [5], Gaussian quadratic [6] and Wyner-Ahlsvede-Körner (WAK) [7], [8] problems. The last setting refers to two correlated sources that are separately compressed, and their compressed versions are conveyed to the decoder, which reproduces only one of the sources in a lossless manner. We consider the WAK problem with conferencing (Fig. 1) in which a pair of correlated sources (X_1^n, X_2^n) are compressed by two encoders that are connected via a one-sided rate-limited link that extends from the 1st encoder to the 2nd. The compressed versions are conveyed to the decoder that outputs an empirical coordination sequence Y^n from which X_1^n can be reproduced in a lossless manner.

Source coordination is an alternative formulation for lossy source coding. *Strong coordination* was considered by Wyner [9], while *empirical coordination* was studied in [10]–[12]. Cuff *et al.* extended these results to the multiuser case [13]. Rather than sending data from one point to another with a fidelity constraint, in a coordination problem all network nodes should develop certain joint statistics. Moreover, it was shown in [13] that rate-distortion theory is a special case of source coordination. In this work, we consider empirical coordination, a problem in which the terminals, upon observing correlated sources, generate sequences with a desired empirical joint distribution. A closely related empirical coordination problem was presented by Berezhi *et al.* [14], who considered a triangular multiterminal network. In this setting, each of the two terminals receives a different correlated source that it compresses and conveys to the decoder. The decoder outputs a sequence that achieves the desired coordination. Moreover, the encoders in [14] may share information via a one-sided cooperation link (see [15] and references therein for additional work involving cooperation in source coding problems). The main contributions of [14] comprise inner and outer bounds on the optimal rate region.

The WAK problem with cooperation considered here is a special case of the triangular multiterminal network in [14] where the sequence X_1^n is losslessly reproduced from the output coordination sequence. We derive a single-letter characterization of the coordination-capacity region for this problem. The direct proof unifies several concepts in source coding by relying on WZ coding [5], binning [16] and superposition coding [17]. Note that in the classical

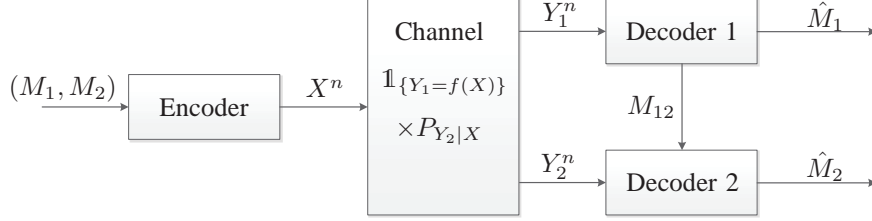


Fig. 2: SD-BC with one-sided decoder cooperation.

WAK problem, where the encoders are non-cooperative, coordination of the output with the side information (i.e., the sequence X_2^n in Fig. 1) is achieved even though it is not required. Therefore, adding such a coordination constraint to the classic WAK problem does not alter its solution, which can be obtained as a special case of the rate region we give here. The non-cooperative version of the problem in Fig. 1, i.e., where one of the sources is losslessly reproduced while coordination with the other source is required, was studied by Berger and Yeung in [18].

To explore duality, we consider a channel coding problem (Fig. 2) that we show is *dual* to the WAK problem of interest. By interchanging the roles of the encoders and decoder of the WAK problem, we obtain a semi-deterministic (SD) broadcast channel (BC) where the decoders cooperate via a rate-limited link. This duality naturally extends the well-known duality between point-to-point (PTP) source and channel coding problems. PTP duality has been widely treated in the literature since it was studied by Shannon in 1959 [19] (see [20]–[22] and references therein). Multiuser duality, however, remains obscure, despite the attention it attracted in the last decade [15], [23]–[25]. We provide principles according to which the two problems can be transformed from one to the other. Moreover, we show that the admissible rate regions of the considered SD-BC and WAK problems are dual. The duality is in the sense that the information measures that define the corner points of both regions coincide, which extends the relation between dual results in the PTP situation.

Cooperative communication over noisy channels was extensively treated in the literature since it was introduced by Willems in the context of a multiple-access channel (MAC), in which the encoders are able to hold a conference [1]. The Gaussian case was solved by Bross *et al.* in [26], followed by several works involving the compound MAC [27], [28]. Cooperation between receivers in a broadcast channel (BC) was introduced by Dabora and Servetto [29]. Liang and Veeravalli generalized the work in [29] by examining the problem of a relay-BC (RBC) [30]. In both [29] and [30], the capacity region of the physically degraded BC (PD-BC) is characterized. Here we combine cooperation in a SD-BC setting.

The SD-BC without cooperation was solved by Gelfand and Pinsker [31]. The coding scheme was based on Marton's scheme for BCs [32] (see [33] for a generalization of [31] to the state-dependent case). We derive the capacity region of the SD-BC with cooperation by first deriving an inner bound on the capacity region of the cooperative general BC. The achievable scheme combines rate-splitting with Marton and superposition coding. The

cooperation protocol uses binning to increase the transmission rate to the cooperation-aided user. The inner bound is then reduced to the SD-BC case and shown to be tight by providing a converse. The presented converse proof takes a simple and compact form by leveraging telescoping identities [34].

There is a close relation between the SD-BC with cooperation and a class of SD-RBCs considered in [35]. We show that a SD-RBC with an orthogonal and deterministic relay is operationally equivalent to the SD-BC with cooperation (see [36] for a related work on equivalence between PTP channels in a general network and noiseless bit-pipes with the same capacity). Consequently, the capacity regions of the two problems are the same. However, there are several advantages of our approach. First, we present a capacity achieving coding scheme over a *single* transmission block, while [35] relies on transmitting many blocks and applying backward decoding. Thus, our scheme avoids the delay introduced by backward decoding. Second, our converse proof is considerably simpler than in [35]. Finally, considering the SD-BC with a one-sided conferencing link between the decoders gives insight into multiuser channel-source duality [37], [38].

To show the duality between the optimal rate regions of the considered source and channel coding problems, an alternative characterization of the capacity region of the SD-BC is given. The corner points of the alternative region satisfy the correspondence to those of the coordination-capacity region of the WAK problem. The structure of the alternative expression motivates a converse proof technique that generalizes classical techniques. Specifically, our converse uses auxiliary random variables that are not only *chosen as a function of the joint distribution induced by each codebook*, but that are *constructed in a probabilistic manner* (see [33] for a deterministic codebook-dependent construction of auxiliaries). Allowing a probabilistic construction of the auxiliary random variables introduces additional optimization parameters (i.e., a probability distribution). By optimizing over the probability values, an upper bound on the alternative formulation of the capacity region is tightened to coincide with the achievable region. Probabilistic arguments of a similar nature were previously used in the literature [39]–[41]. The novelty of our approach, however, is the incorporation of such arguments in a converse proof to describe the optimal choice of auxiliaries. Moreover, a closed form formula for the optimal probability values is derived as part of the converse and highlights the dependence of the choice of auxiliaries on the codebook.

This paper is organized as follows. In Section II we describe the two models of interest - the WAK problem with encoder cooperation and the SD-BC with decoder cooperation. In Section III, we state capacity results for the WAK and BC models. In Section IV we analyse the duality between the two problems and their capacity regions. In Section V we discuss the relation of the considered SD-BC to a class of SD-RBCs. Section VI presents special cases of the capacity region of the SD-BC, and each case is shown to preserve a dual relation to the corresponding reduced source coding problem. Finally, Section VII summarizes the main achievements and insights of this work.

II. PRELIMINARIES AND PROBLEM DEFINITIONS

We use the following notations. Given two real numbers a, b , we denote by $[a:b]$ the set of integers $\{n \in \mathbb{N} \mid [a] \leq n \leq [b]\}$. We define $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Calligraphic letters denote discrete sets, e.g., \mathcal{X} , while the cardinality of a set \mathcal{X} is denoted by $|\mathcal{X}|$. \mathcal{X}^n stands for the n -fold Cartesian product of \mathcal{X} . An element of \mathcal{X}^n is denoted

by $x^n = (x_1, x_2, \dots, x_n)$, and its substrings as $x_i^j = (x_i, x_{i+1}, \dots, x_j)$; when $i = 1$, the subscript is omitted. We define $x^{n \setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., \mathbf{x} . Random variables are denoted by uppercase letters, e.g., X , with similar conventions for random vectors. X_i^j stands for the sequence of random variables $(X_i, X_{i+1}, \dots, X_j)$, while \mathbf{X} stands for X^n . The probability of an event \mathcal{A} is denoted by $\mathbb{P}(\mathcal{A})$, while $\mathbb{P}(\mathcal{A}|\mathcal{B})$ denotes conditional probability of \mathcal{A} given \mathcal{B} . We use $\mathbb{1}_{\mathcal{A}}$ to denote the indicator function of \mathcal{A} . Probability mass functions (PMFs) are denoted by the capital letter P , with a subscript that identifies the random variable and its possible conditioning. For example, for two jointly distributed random variables X and Y , let P_X , $P_{X,Y}$ and $P_{X|Y}$ denote, respectively, the PMF of X , the joint PMF of (X, Y) and the conditional PMF of X given Y . In particular, when X and Y are discrete, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y)$. We omit the subscripts if the arguments of the PMF are lowercase versions of the random variables. If the entries of X^n are drawn in an independent and identically distributed (i.i.d.) manner according to P_X , then for every $\mathbf{x} \in \mathcal{X}^n$ we have $P_{X^n}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n}(\mathbf{x}) = P_X^n(\mathbf{x})$. Similarly, if for every $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ we have $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, we write $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. Finally, for every sequence $\mathbf{x} \in \mathcal{X}^n$, the empirical PMF of \mathbf{x} is defined as follows.

Definition 1 (Empirical PMF) *The empirical PMF of a sequence $\mathbf{x} \in \mathcal{X}^n$ is*

$$\nu_{\mathbf{x}}(a) \triangleq \frac{N(a|\mathbf{x})}{n} \quad (1)$$

where $N(a|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i=a\}}$.

We use $\mathcal{T}_{\epsilon}^{(n)}(P_X)$ to denote the set of letter-typical sequences of length n with respect to the PMF P_X and the non-negative number ϵ [42, Ch. 3], [43], i.e., we have

$$\mathcal{T}_{\epsilon}^{(n)}(P_X) = \left\{ \mathbf{x} \in \mathcal{X}^n : |\nu_{\mathbf{x}}(a) - P_X(a)| \leq \epsilon P_X(a), \forall a \in \mathcal{X} \right\}. \quad (2)$$

Furthermore, for a PMF $P_{X,Y}$ over $\mathcal{X} \times \mathcal{Y}$ and a fixed sequence $\mathbf{y} \in \mathcal{Y}^n$, we define

$$\mathcal{T}_{\epsilon}^{(n)}(P_{X,Y}|\mathbf{y}) = \left\{ \mathbf{x} \in \mathcal{X}^n : (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{\epsilon}^{(n)}(P_{X,Y}) \right\}. \quad (3)$$

A. Wyner-Ahlsvede-Körner Source Coordination Problem with One-Sided Encoder Cooperation

Consider the source coding problem illustrated in Fig. 1. Two source sequences $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ are available at Encoder 1 and Encoder 2, respectively. The sources are drawn in a pairwise independent and identically distributed (i.i.d.) manner according to the PMF P_{X_1, X_2} . Each encoder communicates with the decoder by sending a message via a noiseless communication link of limited rate. The rate of the link between Encoder j and the decoder is R_j and the corresponding message is T_j , where $j = 1, 2$. Moreover, Encoder 1 can communicate with Encoder 2 over a one-sided communication link of rate R_{12} .

Definition 2 (Coordination Code) A $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ coordination code \mathcal{L}_n for the WAK source coordination problem with one-sided encoder cooperation has:

- 1) Three message sets: $\mathcal{T}_{12} = [1 : 2^{nR_{12}}]$, $\mathcal{T}_1 = [1 : 2^{nR_1}]$ and $\mathcal{T}_2 = [1 : 2^{nR_2}]$.
- 2) An encoder cooperation function:

$$f_{12} : \mathcal{X}_1^n \rightarrow \mathcal{T}_{12}. \quad (4a)$$

- 3) Two encoding functions:

$$f_1 : \mathcal{X}_1^n \rightarrow \mathcal{T}_1, \quad (4b)$$

$$f_2 : \mathcal{X}_2^n \times \mathcal{T}_{12} \rightarrow \mathcal{T}_2. \quad (4c)$$

- 4) A decoding function:

$$\phi : \mathcal{T}_1 \times \mathcal{T}_2 \rightarrow \mathcal{Y}^n. \quad (4d)$$

Definition 3 (Total Variation) Let P and Q be two probability measures on the same σ -algebra \mathcal{F} of subsets of the sample space \mathcal{X} . The total variation between P and Q is

$$\|P - Q\|_{TV} = \sup_{\mathcal{A} \in \mathcal{F}} |P(\mathcal{A}) - Q(\mathcal{A})|. \quad (5)$$

Remark 1 If the sample space \mathcal{X} is countable, the total variation between P and Q reduces to ¹

$$\|P - Q\|_{TV} = \frac{1}{2} \sum_{a \in \mathcal{X}} |P(a) - Q(a)|. \quad (6)$$

Let \mathcal{P} be a set of PMFs $P_{X_1, X_2, Y}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ that factor as $P_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1=f(Y)\}}$ and have P_{X_1, X_2} as a marginal.

Definition 4 (Coordination Achievability) Let $P_{X_1, X_2, Y} \in \mathcal{P}$. A rate triple (R_{12}, R_1, R_2) is $P_{X_1, X_2, Y}$ -achievable if for every $\epsilon, \delta > 0$ there is a sufficiently large $n \in \mathbb{N}$ and a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ coordination code \mathcal{L}_n such that ²

$$\mathbb{P} \left(\left\| \nu_{X_1^n, X_2^n, Y^n} - P_{X_1, X_2, Y} \right\|_{TV} \geq \delta \middle| \mathcal{L}_n \right) \leq \epsilon. \quad (7)$$

Definition 5 (Coordination-Capacity Region) The coordination-capacity region with respect to a PMF $P_{X_1, X_2, Y} \in \mathcal{P}$ is the closure of the set of rate triples (R_{12}, R_1, R_2) that are $P_{X_1, X_2, Y}$ -achievable.

B. The SD-BC Channel with One-Sided Decoder Cooperation

The SD-BC with cooperation is illustrated in Fig. 2. The channel has one sender and two receivers. The sender chooses a pair (m_1, m_2) of indices uniformly and independently from the set $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ and maps

¹Countable sample spaces are assumed throughout this work

²Being clear from the context, in this work we drop the conditioning on \mathcal{L}_n .

them to a sequence $\mathbf{x} \in \mathcal{X}^n$, which is the channel input. The sequence \mathbf{x} is transmitted over a BC with transition probability $P_{Y_1, Y_2 | X} = \mathbb{1}_{\{Y_1 = f(X)\}} P_{Y_2 | X}$. The output sequence $\mathbf{y}_j \in \mathcal{Y}_j^n$, where $j = 1, 2$, is received by decoder j . Decoder j produces an estimate of m_j , which is denoted by \hat{m}_j . There is a one-sided noiseless cooperation link of rate R_{12} from Decoder 1 to Decoder 2. By conveying a message $m_{12} \in [1 : 2^{nR_{12}}]$ over this link, Decoder 1 can share with Decoder 2 information about \mathbf{y}_1 , \hat{m}_1 , or both.

Definition 6 (Code Description) A $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code \mathcal{C}_n for the SD-BC with one-sided decoder cooperation has:

1) Three message sets $\mathcal{M}_{12} = [1 : 2^{nR_{12}}]$, $\mathcal{M}_1 = [1 : 2^{nR_1}]$ and $\mathcal{M}_2 = [1 : 2^{nR_2}]$.

2) An encoding function:

$$g : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n. \quad (8a)$$

3) A decoder cooperation function:

$$g_{12} : \mathcal{Y}_1^n \rightarrow \mathcal{M}_{12}. \quad (8b)$$

4) Two decoding functions:

$$\psi_1 : \mathcal{Y}_1^n \rightarrow \mathcal{M}_1, \quad (8c)$$

$$\psi_2 : \mathcal{Y}_2^n \times \mathcal{M}_{12} \rightarrow \mathcal{M}_2. \quad (8d)$$

Definition 7 (Error Probability) The average error probability for the $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code \mathcal{C}_n is

$$P_e(\mathcal{C}_n) = \mathbb{P}\left((\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \mid \mathcal{C}_n\right) \quad (9a)$$

where $\hat{M}_1 = \psi_1(Y_1^n)$ and $\hat{M}_2 = \psi_2(Y_2^n, M_{12})$. The average error probability for each receiver is defined by

$$P_{e,1}(\mathcal{C}_n) = \mathbb{P}\left(\hat{M}_1 \neq M_1 \mid \mathcal{C}_n\right) ; P_{e,2}(\mathcal{C}_n) = \mathbb{P}\left(\hat{M}_2 \neq M_2 \mid \mathcal{C}_n\right). \quad (9b)$$

Definition 8 (Achievable Rates) A rate triple (R_{12}, R_1, R_2) is achievable if for any $\epsilon > 0$ there is a sufficiently large $n \in \mathbb{N}$ and a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code \mathcal{C}_n such that $P_e(\mathcal{C}_n) \leq \epsilon$.³

The *capacity region* is the closure of the set of achievable rates.

III. MAIN RESULTS

We state our main results as the coordination-capacity region of the WAK source coordination problem (Section II-A) and the capacity region of the SD-BC with cooperation (Section II-B).

³As stated in Section II-A, the conditioning on \mathcal{C}_n will also be dropped subsequently. Accordingly, $P_e(\mathcal{C}_n)$ will be denoted by $P_e^{(n)}$.

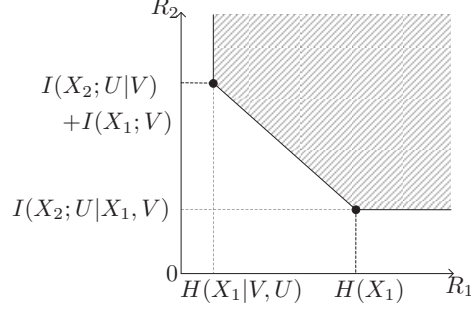


Fig. 3: Corner points of the coordination-capacity region of the WAK coordination problem with cooperation at the hyperplane where $R_{12} = I(X_1; V|X_2)$.

Theorem 2 (Coordination-Capacity Region of the WAK Problem) *The coordination-capacity region with respect to a PMF $P_{X_1, X_2, Y} \in \mathcal{P}$ of the WAK source coordination problem with one-sided encoder cooperation is denoted by \mathcal{R}_{WAK} and is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_{12} \geq I(V; X_1|X_2) \quad (10a)$$

$$R_1 \geq H(X_1|V, U) \quad (10b)$$

$$R_2 \geq I(U; X_2|X_1, V) \quad (10c)$$

$$R_1 + R_2 \geq H(X_1|V, U) + I(V, U; X_1, X_2) \quad (10d)$$

where the union is over all PMFs $P_{X_1, X_2} P_{V|X_1} P_{U|X_2, V} P_{Y|X_1, U, V}$ that have $P_{X_1, X_2, Y}$ as a marginal. Moreover, \mathcal{R}_{WAK} is convex and one may choose $|\mathcal{V}| \leq |\mathcal{X}_1| + 4$ and $|\mathcal{U}| \leq |\mathcal{V}| \cdot |\mathcal{X}_2| + 3$.

See Appendix A for the proof of Theorem 2.

Remark 3 *For a fixed PMF in Theorem 2, the triples (R_{12}, R_1, R_2) at the corner points of \mathcal{R}_{WAK} are (see Fig. 3)*

$$(I(V; X_1|X_2), H(X_1), I(U; X_2|X_1, V)) \quad (11a)$$

$$(I(V; X_1|X_2), H(X_1|V, U), I(U; X_2|V) + I(V; X_1)). \quad (11b)$$

The corner point in (11b) can be achieved using the coding scheme in [14] by setting $V = 0$. However, the rate triple (11a) does not seem to be achievable for that scheme.

Remark 4 *The source coordination problem defined in Section II-A can be transformed into an equivalent rate-distortion problem. This is done by substituting \mathbf{Y} , the output of the coordination problem, with the pair $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$, where $\hat{\mathbf{X}}_1$ is a lossless reconstruction of the source sequence \mathbf{X}_1 , while $\hat{\mathbf{X}}_2$ satisfies the distortion constraint*

$$\mathbb{E} \left[\sum_{i=1}^n d(X_{2,i}, \hat{X}_{2,i}) \right] \leq D \quad (12)$$

where $d : \mathcal{X}_2 \times \hat{\mathcal{X}}_2 \rightarrow \mathbb{R}_+$ is a single-letter distortion measure and $D \in \mathbb{R}_+$ is the distortion constraint. The two models are equivalent in the sense that the rate bounds that describe the optimal rate regions of both problems are the same; the domain over which the union is taken, however, is slightly modified. This gives rise to the following corollary.

Corollary 5 (Rate-Distortion Region of the WAK Problem) *The rate-distortion region for the equivalent rate-distortion problem is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying (10), where the union is over all PMFs $P_{X_1, X_2} P_{V|X_1} P_{U|X_2, V}$ and the reconstructions \hat{X}_2 that are deterministic functions of (X_1, U, V) such that $\mathbb{E}[d(X_2, \hat{X}_2)] \leq D$.*

The proof of Corollary 5 is similar to that of Theorem 2 and is omitted. We next state the capacity region of the SD-BC with cooperation.

Theorem 6 (Capacity Region of the SD-BC) *The capacity region \mathcal{C}_{BC} of the SD-BC with one-sided encoder cooperation is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_1 \leq H(Y_1) \tag{13a}$$

$$R_2 \leq I(V, U; Y_2) + R_{12} \tag{13b}$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) \tag{13c}$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(V, U; Y_2) + R_{12} \tag{13d}$$

where the union is over all PMFs $P_{V, U, Y_1} P_{X|V, U, Y_1} P_{Y_2|X}$ for which $Y_1 = f(X)$. Moreover, \mathcal{C}_{BC} is convex and one may choose $|\mathcal{V}| \leq |\mathcal{X}| + 3$ and $|\mathcal{U}| \leq |\mathcal{V}| \cdot |\mathcal{X}| + 1$.

The proof of Theorem 6 is relegated to Appendix B. The achievable scheme combines Marton and superposition coding with rate-splitting and binning. The rather simple converse proof is due to the telescoping identity [34, Eq. (9) and (11)].

Remark 7 *The derivation of the capacity region in Theorem 6 strongly relies on the SD nature of the channel. Since $Y_1 = f(X)$, the encoder has full control over the message that is conveyed via the cooperation link. This allows one to design the cooperation protocol at the encoding stage without assuming a particular Markov relation on the coding random variables. Our approach differs from the one taken in [29], where an inner bound on the capacity region of a BC with two-sided conferencing links between the decoders was derived. In [29], the decoders cooperate by conveying compressed versions of their received channel outputs to each other (via a WZ-like coding mechanism). Doing so forced the authors to restrict their coding PMF to satisfy certain Markov relations that must not hold in general. Consequently, the inner bound in [29] is not tight for the SD-BC we consider here.*

Remark 8 *The SD-BC with cooperation is strongly related to the SD-RBC that was studied in [35]. The SD-BC*

with cooperation is operationally equivalent to a reduced version of the SD-RBC, in which the relay channel is orthogonal and deterministic. A full discussion of the relation between the settings is given in Section V.

Remark 9 The cardinality bounds on the auxiliary random variables V and U in Theorems 2 and 6 are established by invoking the support lemma [44, Lemma 15.4] twice. The details are omitted.

Remark 10 The SD-BC with decoder cooperation and the WAK problem with encoder cooperation are duals. A full discussion on the duality between the problems is given in the following section.

IV. DUALITY BETWEEN THE SOURCE AND CHANNEL CODING PROBLEMS

We examine the WAK coordination problem with encoder cooperation (Fig. 1) and the SD-BC with decoder cooperation (Fig. 2) from a duality perspective. We show that the two problems and their solutions are dual to one another in a manner that naturally extends PTP duality [20]–[22]. In the PTP scenario, two lossy source (or, equivalently, source coordination) and channel coding problems are said to be dual if interchanging the roles of the encoder and the decoder in one problem produces the other problem. The solutions of these problems are dual in that they require an optimization of an information measure of the same structure, up to renaming the random variables involved. Solving one problem provides insight into the solution of the other. However, how duality extends to the multiuser case is still obscure.

In the context of multiuser lossy source coding, we favor the framework of source coordination over rate-distortion, since the former provides a natural perspective on the similarities of the two problems. Source coordination inherently accounts for the probabilistic relations among *all* the sequences involved in the problem's definition. However, in a coordination problem, both the input and output (coordination) PMFs are fixed, while in a channel coding problem, the input PMF is optimized. Therefore, for convenience, throughout this section we consider channel codes with codewords of a fixed type, as defined in the following (see also [15]).

Definition 9 (Fixed-Type Code Description, Achievability and Capacity Region) A

$(2^{nR_{12}}, 2^{nR_1}, 2^{nR_2}, n, P_X^*)$ fixed-type code \mathcal{C}_n^* for the SD-BC with one-sided decoder cooperation consists of three integer sets, an encoding function, a decoder cooperation function, and two decoding functions as defined in (8). For any $n \in \mathbb{N}$ and $\delta > 0$, we define the error probability of a code \mathcal{C}_n^* as

$$P_e^{(\delta)}(\mathcal{C}_n^*) = \mathbb{P}\left(\left\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\right\} \cup \left\{\left\|\nu_{X^n, Y_1^n, Y_2^n} - P_X^* \mathbb{1}_{\{Y_1=f(X)\}} P_{Y_2|X}\right\|_{TV} \geq \delta\right\} \middle| \mathcal{C}_n^*\right). \quad (14)$$

The error probability for each receiver is defined analogously. A rate triple (R_{12}, R_1, R_2) is achievable if for any $\epsilon, \delta > 0$, there is a sufficiently large $n \in \mathbb{N}$ and a $(2^{nR_{12}}, 2^{nR_1}, 2^{nR_2}, n, P_X^*)$ fixed-type code \mathcal{C}_n^* such that $P_e^{(\delta)}(\mathcal{C}_n^*) \leq \epsilon$. The definition of the corresponding capacity region is standard (see [45]).

Note that fixed-composition codes [46]–[49] and codes that are drawn i.i.d. according to P_X^* satisfy the property in Definition 9. Moreover, the capacity region of the SD-BC with cooperation and a fixed-type code is similar to

TABLE I: Duality transformation principles: the WAK problem with cooperation vs. the SD-BC with cooperation

WAK Problem with Encoder Cooperation	Semi-Deterministic BC with Decoder Cooperation
Decoder inputs / Encoder outputs: $T_j \in [1 : 2^{nR_j}], j = 1, 2$	Encoder inputs / Decoder outputs: $M_j \in [1 : 2^{nR_j}], j = 1, 2$
Encoder inputs / Sources: $\mathbf{X}_1, \mathbf{X}_2$	Decoder inputs / Channel outputs: $\mathbf{Y}_1, \mathbf{Y}_2$
Decoder output / Coordination sequence: \mathbf{Y}	Encoder output / Channel input: \mathbf{X}
Encoding functions: $f_1 : \mathcal{X}_1^n \rightarrow \mathcal{T}_1,$ $f_2 : \mathcal{X}_2^n \times \mathcal{T}_{12} \rightarrow \mathcal{T}_2$	Decoding functions: $\psi_1 : \mathcal{Y}_1^n \rightarrow \mathcal{M}_1,$ $\psi_2 : \mathcal{Y}_2^n \times \mathcal{M}_{12} \rightarrow \mathcal{M}_2$
Encoder cooperation functions: $f_{12} : \mathcal{X}_1^n \rightarrow \mathcal{T}_{12}$	Decoder cooperation function: $g_{12} : \mathcal{Y}_1^n \rightarrow \mathcal{M}_{12}$
Decoding functions: $\phi : \mathcal{T}_1 \times \mathcal{T}_2 \rightarrow \mathcal{Y}^n$	Encoding function: $g : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n$

that stated in Theorem 6. The only difference between the regions is the domain of PMFs over which the union is taken. Specifically, for the BC with a fixed-type code, the union is taken over all PMFs $P_{V,U,Y_1} P_{X|V,U,Y_1} P_{Y_2|X}$ that have $P_X^* \mathbb{1}_{\{Y_1=f(X)\}} P_{Y_2|X}$ as a marginal.

The WAK and SD-BC problems with cooperation are obtained from the other by interchanging the roles of their encoder(s) and decoder(s) and renaming the random variables involved. A full description of the duality transformation principles is given in Table I. The duality is also evident in that the input and output sequences in both problems are jointly typical with respect to a PMF of the same form. Namely, in the source coding problem, the triple $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ is coordinated with respect to the PMF

$$P_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1=f(Y)\}} = P_Y \mathbb{1}_{\{X_1=f(Y)\}} P_{X_2|Y}. \quad (15)$$

The corresponding triple of sequences $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2)$ in the channel coding problem are jointly typical with high probability with respect to the PMF

$$P_X^* \mathbb{1}_{\{Y_1=f(X)\}} P_{Y_2|X}. \quad (16)$$

By renaming the random variables according to Table I, the two PMFs in (15) and (16) coincide.

The duality between the two problems extends beyond the correspondence presented above. The coordination-capacity region of the WAK problem (Theorem 2) and the capacity region of the BC (Theorem 6) are also dual to one another. To see this, the following lemma gives an alternative characterization of the capacity region C_{BC} .

Lemma 11 (Alternative Characterization of SD-BC Capacity) *Let $\mathcal{C}_{BC}^{(D)}$ be the region defined by the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_{12} \geq I(V; Y_1) - I(V; Y_2) \quad (17a)$$

$$R_1 \leq H(Y_1) \quad (17b)$$

$$R_2 \leq I(V, U; Y_2) + R_{12} \quad (17c)$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) \quad (17d)$$

where the union is over the same domain stated in Theorem 6. Then:

$$\mathcal{C}_{BC}^{(D)} = \mathcal{C}_{BC}. \quad (18)$$

See Appendix D for a proof of Lemma 11 based on bidirectional inclusion arguments.

Remark 12 $\mathcal{C}_{BC}^{(D)}$ can be established as the capacity region of the SD-BC with cooperation by providing achievability and converse proofs. We refer the reader to [50] for a full description of the achievability scheme. The proof of the converse is given in Appendix E. The converse is established via a novel approach, in which the auxiliaries are not only chosen as a function of the joint distribution induced by each codebook, but they are also constructed in a probabilistic manner. The need for this probabilistic construction stems from the unique structure of the region $\mathcal{C}_{SD}^{(D)}$. Specifically, the lower bound on R_{12} in (17a) (which is typical to source coding problems where the random source sequences are commonly memoryless) and the fact that Y_1^n and Y_2^n have memory are the underlying reasons for the usefulness of the approach. Depending on the distribution that stems from the codebook, a deterministic choice of auxiliaries may result in a $I(V; Y_1) - I(V; Y_2)$ that is too large. By a stochastic choice of the auxiliaries, we circumvent this difficulty and dominate the quantity $I(V; Y_1) - I(V; Y_2)$ to satisfy (17a).

The converse proof boils down to two key steps. First, we derive an outer bound on the achievable region $\mathcal{C}_{BC}^{(D)}$ that is described by three auxiliary random variables (A, B, C) . Then, by probabilistically choosing (V, U) from (A, B, C) , we show that the outer bound is tight. The second step implies that the outer bound is an alternative formulation of the capacity region. Capacity proofs that rely on alternative descriptions for which the converse is provided have been used previously (see, e.g., [51] and [52]). However, the proof of equivalence typically relies on operational arguments rather than on a probabilistic identification of auxiliaries. Probabilistic arguments of a similar nature to those we present here were also used before [39]–[41]. For instance, in [39], probabilistic arguments were used to prove the equivalence between two representations of the compress-and-forward inner bound for the relay channel via time-sharing. Such arguments were also leveraged in [40] to characterize the admissible rate-distortion region for the multiterminal source coding problem under logarithmic loss. The novelty of our approach stems from combining these two concepts and essentially using a probabilistic construction to define the auxiliary random variables and establish the tightness of the outer bound. We derive a closed form formula for the optimal probability values, that highlights the dependence of the the auxiliaries on the distribution induced by the codebook.

The duality between \mathcal{R}_{WAK} in (10) and $\mathcal{C}_{BC}^{(D)}$ in (17) is expressed in a correspondence between the information measures at their corner points. The values of (R_{12}, R_1, R_2) at the corner points of the coordination-capacity region

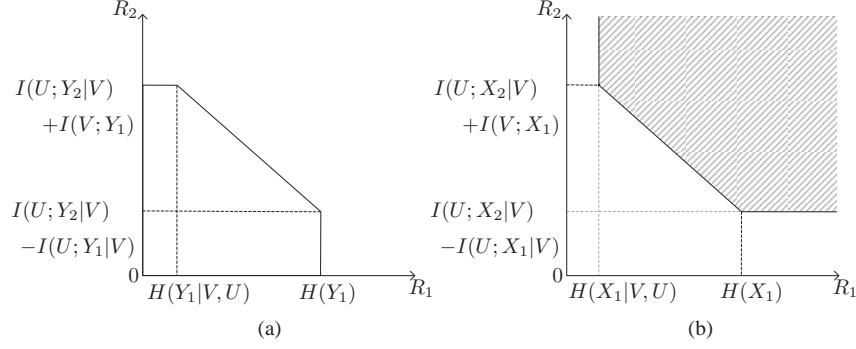


Fig. 4: Corner point correspondence between: (a) the capacity region of the SD-BC with cooperation; (b) the coordination-capacity region of the WAK coordination problem with cooperation. The regions are depicted at the hyperplanes where to $R_{12} = I(V; Y_1) - I(V; Y_2)$ and $R_{12} = I(V; X_1) - I(V; X_2)$, respectively.

of the WAK problem are

$$(I(V; X_1|X_2), H(X_1), I(U; X_2|X_1, V)) \quad (19a)$$

$$(I(V; X_1|X_2), H(X_1|V, U), I(U; X_2|V) + I(V; X_1)) \quad (19b)$$

while the triple (R_{12}, R_1, R_2) at the corner points of capacity region of the SD-BC with cooperation are

$$(I(V; Y_1) - I(V; Y_2), H(Y_1), I(U; Y_2|V) - I(U; Y_1|V)) \quad (20a)$$

$$(I(V; Y_1) - I(V; Y_2), H(Y_1|V, U), I(U; Y_2|V) + I(V; Y_1)). \quad (20b)$$

We show that (19a)-(19b) and (20a)-(20b) correspond by first rewriting the value of R_{12} in (19a)-(19b) as

$$R_{12} = I(V; X_1|X_2) \stackrel{(a)}{=} I(V; X_1) - I(V; X_2) \quad (21)$$

where (a) follows from Markov relation $V - X_1 - X_2$. Moreover, the value of R_2 in (19a) is rewritten as

$$R_2 = I(U; X_2|X_1, V) \stackrel{(a)}{=} I(U; X_2|V) - I(U; X_1|V) \quad (22)$$

where (a) follows from the Markov relation $U - (X_2, V) - X_1$. By substituting (21)-(22) into (19a)-(19b) and renaming the random variables according to Table I, the corner points of both regions coincide (see Fig. 4).

Chronologically, upon observing the duality between the two problem settings, we solved the WAK problem first. Then, based on past experience (cf., e.g., [15] and [25]), our focus turned to the dual SD-BC with cooperation. Since the capacity region is defined by the corner points of a union of polytopos, the structure of the capacity region for the SD-BC was evident. Thus, duality was key in obtaining the results of this work. We note that the relation between our result for the SD-BC with cooperation and the SD-RBC (that is discussed in the following section) was observed only at a later stage.

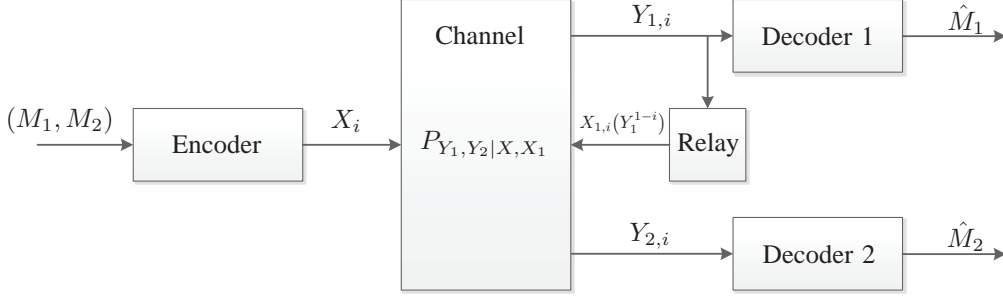


Fig. 5: General RBC.

V. RELATION TO THE SEMI-DETERMINISTIC RELAY BROADCAST CHANNEL

The SD-BC with cooperation is strongly related to the SD-RBC that was studied in [35]. A general RBC is illustrated in Fig. 5 (for the full definition see [35, Section II]). The RBC is SD if the PMF $P_{Y_1|X, X_1}$ only takes on the values 0 or 1. To see the correspondence between the SD-RBC and the BC of interest, let $Y_2 = (Y_{21}, Y_{22})$ and let the channel transition PMF factorize as

$$P_{Y_1, Y_{21}, Y_{22}|X, X_1} = P_{Y_{21}|X} \mathbb{1}_{\{Y_1=f(X)\}} P_{Y_{22}|X_1}. \quad (23)$$

(23) implies that the channel from the encoder to the decoders is orthogonal to the one between the decoders. Suppose the relay channel is deterministic with capacity R_{12} and let $Y_{22} = f_R(X_1)$. The SD-RBC obtained under these assumptions is referred to as the R_{12} -reduced SD-RBC and its capacity region is denoted by $\mathcal{C}_{RBC}(R_{12})$. As stated in the following lemma, the R_{12} -reduced SD-RBC is operationally equivalent to the SD-BC with cooperation. By operational equivalence, we mean that for every achievable rate tuple in one problem, there exists a code (that achieves these rates) that can be transformed into a code (with the same rates) for the other problem. The transformation mechanism treats the code for each model as a black box and is described as part of the proof of Lemma 13.

Lemma 13 (Operational Equivalence) *For every $(R_1, R_2) \in \mathcal{C}_{RBC}(R_{12})$, there is a $(n, 2^{nR_1}, 2^{nR_2})$ code $\mathcal{C}_n^{(RBC)}(R_{12})$ for the R_{12} -reduced SD-RBC that can be transformed into a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code $\mathcal{C}_n^{(BC)}$ for the SD-BC with cooperation, and vice versa. Namely, for every $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}$, there is a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code $\mathcal{C}_n^{(BC)}$ for the SD-BC with cooperation that can be transformed into a $(n, 2^{nR_1}, 2^{nR_2})$ code $\mathcal{C}_n^{(RBC)}(R_{12})$ for the R_{12} -reduced SD-RBC.*

Proof: Fix $(R_1, R_2) \in \mathcal{C}_{RBC}(R_{12})$ and let $\left\{ \mathcal{C}_{RBC}^{(n)}(R_{12}) \right\}_{n=1}^{\infty}$ be the sequence of $(n, 2^{nR_1}, 2^{nR_2})$ codes for the R_{12} -reduced SD-RBC that adhere to the coding scheme described in [35, Appendix I]. Accordingly, $P_e(\mathcal{C}_{RBC}^{(n)}(R_{12})) \rightarrow 0$ as $n \rightarrow \infty$ and the induced codewords, channel inputs, and channel outputs are jointly-typical

with high probability ⁴. Since the channel from Decoder 1 to Decoder 2 is deterministic, there are approximately $2^{nH(Y_{22})}$ different possible relay channel outputs \mathbf{y}_{22} . Recall that the capacity of the orthogonal and deterministic relay of the R_{12} -reduced SD-RBC is exactly R_{12} , i.e., $H(Y_{22}) = R_{12}$. For every sequence $\mathbf{y}_{22} \in \mathcal{T}_\epsilon^{(n)}(P_{Y_{22}})$ (here $\epsilon > 0$ corresponds to the margin between the region achieved by the n th code in the sequence and (R_1, R_2)), define the following subset of \mathbf{x}_1 codewords:

$$\mathcal{V}(\mathbf{y}_{22}) = \left\{ \mathbf{x}_1 \in \mathcal{C}_{RBC}^{(n)}(R_{12}) \mid f_R(x_{1,i}) = y_{22,i}, \forall i \in [1 : n] \right\}. \quad (24)$$

Consider a SD-BC with cooperation and associate a cooperation message m_{12} , where $m_{12} \in \mathcal{M}_{12}$, with every set $\mathcal{V}(\mathbf{y}_{22})$. To use $\mathcal{C}_{RBC}^{(n)}$ for the SD-BC with cooperation, Decoder 1 waits for the n -symbol transmission to end and then shares with Decoder 2 the message m_{12} associated with a set $\mathcal{V}(\mathbf{y}_{22})$ that contains the intended \mathbf{x}_1 codeword (i.e., such that $\mathbf{x}_1 \in \mathcal{V}(\mathbf{y}_{22})$). Given m_{12} , Decoder 2 recovers the sequence \mathbf{y}_{22} and proceeds with the decoding process of the R_{12} -reduced SD-RBC coding scheme. This results in a sequence of $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ codes $\left\{ \mathcal{C}_n^{(BC)} \right\}_{i=1}^\infty$ for the SD-BC with cooperation.

Next, fix $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}$ and let $\left\{ \mathcal{C}_{BC}^{(n)} \right\}_{n=1}^\infty$ be the sequence of $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ codes for the SD-BC with cooperation described in Appendix B. Consider an R_{12} -reduced SD-RBC and map each cooperation message $m_{12} \in \mathcal{M}_{12}$ to a codeword $\mathbf{x}_1(m_{12})$. Since the capacity of the channel between the decoders is R_{12} , there is a sufficient number of different codewords \mathbf{x}_1 (i.e., sufficient to cover the space of cooperation messages $\mathcal{M}_{12} = [1 : 2^{nR_{12}}]$) that can be conveyed via this channel. To use $\mathcal{C}_{BC}^{(n)}$ for the R_{12} -reduced SD-RBC, we transmit B blocks, each of length n , and denote the messages transmitted by $(m_1^{(b)}, m_2^{(b)}) \in \mathcal{M}_1 \times \mathcal{M}_2$, where $b \in [1 : B]$. In the subsequent coding scheme, the transmission of the 1st block is disregarded, while during every block $b \geq 2$, the messages $(m_1^{(b-1)}, m_2^{(b-1)})$ are reliably transmitted over the channel. Accordingly, the scheme forfeits the decoding of the messages $(m_1^{(B)}, m_2^{(B)})$, which implies that the average rate pair $(\frac{B-1}{B}R_1, \frac{B-1}{B}R_2)$, over B blocks, is achievable. By taking $B \rightarrow \infty$, the transmission rates approach (R_1, R_2) .

The coding scheme for the R_{12} -reduced SD-RBC during block $b \geq 2$ is described subsequently. First note that the channel output $\mathbf{y}_1^{(b-1)}$ at Decoder 1 during the previous block is known at the relay at the beginning of block b . Thus, during block b , the encoder transmits the codeword \mathbf{x} that corresponds to the message pair $(m_1^{(b-1)}, m_2^{(b-1)})$, while the relay transmits the codeword $\mathbf{x}_1 \left(m_{12}^{(b-1)}(\mathbf{y}_1^{(b-1)}) \right)$. At the end of transmission block b , Decoder 2 uses the induced relay output $\mathbf{y}_{22}^{(b)}$ to reliably decode $m_{12}^{(b-1)}(\mathbf{y}_1^{(b-1)})$. Both decoders then proceed with the decoding process for the SD-BC with cooperation to decode the messages $(m_1^{(b-1)}, m_2^{(b-1)})$. By taking n and B to infinity, this coding scheme achieves (R_1, R_2) for the R_{12} -reduced SD-RBC. ■

Lemma 13 implies that the capacity regions of the SD-BC with cooperation and the R_{12} -reduced SD-RBC coincide. Using the result of [35, Theorem 8], the capacity region of the R_{12} -reduced SD-RBC is the union of rate

⁴ $P_e(\mathcal{C})$ stands for the error probability induced by a code \mathcal{C}

pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:

$$R_1 \leq H(Y_1|X_1) \quad (25a)$$

$$R_2 \leq I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21}) \quad (25b)$$

$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(U; Y_{21}|V, X_1) + I(V; Y_1|X_1) \quad (25c)$$

$$R_1 + R_2 \leq H(Y_1|V, U, X_1) + I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21}) \quad (25d)$$

where the union is over all PMFs $P_{V,U,X,X_1} P_{Y_{21}|X} \mathbb{1}_{\{Y_1=f(X)\}} \mathbb{1}_{\{Y_{22}=f_R(X_1)\}}$. In Appendix F we show that the region in (25) can be simplified to coincide with the capacity region of the SD-BC with cooperation given in Theorem 6.

The advantage of the approach taken in this work compared to that in [35] is threefold. First, we achieve capacity over a *single* transmission block, while the scheme in [35] (which can also be used for the SD-BC with cooperation due to Lemma 13) transmits a large number of blocks and applies backward decoding. The substantial delay introduced by a backward decoding process implies the superiority of our scheme for practical uses. The reduction of the multi-block coding scheme in [35] to our single-block scheme is consistent with the results in [53]. The authors of [53] showed that for the primitive relay channel (i.e., a relay channel with a noiseless link from relay to the receiver), the decode-and-forward and compress-and-forward multi-block coding schemes can be applied with only a single transmission block. The second advantage of our approach is the simple and concise converse proof that follows using telescoping identities [34, Eq. (9) and (11)]. Finally, considering the SD-BC with cooperation (rather than the SD-RBC) is highly applicable for the discussion on duality due to its correspondence with the cooperative WAK source coordination problem. This duality is presented and discussed in Section IV, and it yields insight into the relation between multiuser channel and source coding problems.

VI. SPECIAL CASES

We consider special cases of the capacity region of the SD-BC with decoder cooperation and show that the dual relation discussed in Section IV is preserved for each special case.

A. The Deterministic Broadcast Channel with Decoder Cooperation

Corollary 14 (Capacity Region of the Deterministic BC) *The capacity region of a deterministic BC (DBC) is the union of the set of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_1 \leq H(Y_1) \quad (26a)$$

$$R_2 \leq H(Y_2) + R_{12} \quad (26b)$$

$$R_1 + R_2 \leq H(Y_1, Y_2) \quad (26c)$$

where the union is over all input PMFs P_X .

Proof: Achievability follows from Theorem 6 by taking $V = 0$ and $U = Y_2$. A converse follows by the Cut-Set bound. \blacksquare

The DBC is dual to the SW source coding problem with one-sided encoder cooperation (see [54] and [55]). By inspecting the optimal rate regions for these two problems, we have the duality discussed in Section IV. The SW setting is obtained from the WAK coordination problem by also adding a lossless reproduction requirement to the second source. By properly choosing the auxiliary random variables, the \mathcal{R}_{WAK} reduces to the optimal rate region for the SW problem, which is the set of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \geq H(X_1|X_2) - R_{12} \quad (27a)$$

$$R_2 \geq H(X_2|X_1) \quad (27b)$$

$$R_1 + R_2 \geq H(X_1, X_2) \quad (27c)$$

(see Appendix G for the derivation of (27)), while the capacity region of the DBC is given by (26). Examining the regions in (26) and (27) reveals the correspondence between their corner points.

B. The Physically Degraded Broadcast Channel with Decoder Cooperation

Corollary 15 (Capacity Region of the PD-BC) *The capacity region for the physically degraded case in which $Y_1 = X$ coincides with the results in [29] and [41] and is the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:*

$$R_1 \leq H(X|U) \quad (28a)$$

$$R_2 \leq I(U; Y_2) + R_{12} \quad (28b)$$

$$R_1 + R_2 \leq H(X) \quad (28c)$$

where the union is over all PMFs $P_{U,X}P_{Y_2|X}$.

Proof: The capacity region of the PD-BC was originally derived in [29] where it was described as the union of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(X; Y_1|U) \quad (29a)$$

$$R_2 \leq I(U; Y_2) + R_{12} \quad (29b)$$

$$R_2 \leq I(U; Y_1) \quad (29c)$$

where the union is over all PMFs $P_{U,X}P_{Y_1|X}P_{Y_2|Y_1}$.

An equivalent characterization of region in (29) was later given in [41] as the union over the domain stated above of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(X; Y_1|U) \quad (30a)$$

$$R_2 \leq I(U; Y_2) + R_{12} \quad (30b)$$

$$R_1 + R_2 \leq I(X; Y_1). \quad (30c)$$

Since a SD-BC in which $Y_1 = X$ is also physically degraded, substituting $Y_1 = X$ into (30) yields the region in Corollary 15. By substituting $Y_1 = X$, setting $U = 0$, and relabeling V as U in the capacity of the SD-BC with cooperation stated in Theorem 6, we obtain an achievable region that is the union over the domain stated in Corollary 15 of rate triples $(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_2 \leq I(U; Y_2) + R_{12} \quad (31a)$$

$$R_1 + R_2 \leq H(X). \quad (31b)$$

Denote the capacity region in (28) by \mathcal{C}_{PD} and the achievable region in (31) by \mathcal{R}_{SD} . Since \mathcal{R}_{SD} is an achievable region, clearly $\mathcal{R}_{SD} \subseteq \mathcal{C}_{PD}$. On the other hand, the opposite inclusion $\mathcal{C}_{PD} \subseteq \mathcal{R}_{SD}$ also holds, because the rate bound (28a) does not appear in \mathcal{R}_{SD} , while (28b)-(28c) and the domain over which the union is taken are preserved. ■

The dual source coding problem for the PD-BC with cooperation where $Y_1 = X$, is a model in which the output sequence is a lossless reproduction of \mathbf{X}_1 . The latter setting is a special case of the WAK problem with cooperation, that is obtained by taking f (the coordination function) to be the identity function. The corresponding coordination-capacity region is given by (10) (with a slight modification of the domain over which the union is taken). However, an equivalent coordination-capacity region that is characterized by a single auxiliary random variable has yet to be derived. Since the capacity region of the PD-BC with cooperation where $Y_1 = X$ is described using a single auxiliary (as in (28)), the lack of such characterization for the region of the dual problem makes the comparison problematic. Nonetheless, recalling that the capacity region of the considered PD-BC is also given by (13) while substituting $Y_1 = X$ emphasizes that the duality holds.

VII. SUMMARY AND CONCLUDING REMARKS

We considered the WAK empirical coordination problem with one-sided encoder cooperation and derived its coordination-capacity region. The capacity achieving coding scheme combined WZ coding, binning and superposition coding. Furthermore, a SD-BC in which the decoders can cooperate via a one-sided rate-limited link was considered and its capacity region was found. Achievability was established by deriving an inner bound on the capacity region of a general BC that was shown to be tight for the SD scenario. The coding strategy that achieved the inner bound combined rate-splitting, Marton and superposition coding, and binning (used for the cooperation protocol). The converse for the SD case leveraged telescoping identities that resulted in a concise and a simple proof. The relation between the SD-BC with cooperation and the SD-RBC was examined. The two problems were shown to be operationally equivalent under proper assumptions and the correspondence between their capacity regions was established.

The cooperative WAK and SD-BC problems were inspected from a channel-source duality perspective. Transformation principles between the two settings that naturally extend duality relations between PTP models were

presented. It was shown that the duality between the WAK and the SD-BC problems induces a duality between their capacities that is expressed in a correspondence between the corner points of the two regions. To this end, the capacity region of the SD-BC was restated as an alternative expression. The converse was based on a novel approach where the construction of the auxiliary random variables is probabilistic and depends on the distribution induced by the codebook. The probabilistic construction introduced additional optimization parameters (the probability values) that were used to tighten the outer bound to coincide with the alternative achievable region. To conclude the discussion, several special cases of the BC setting and their corresponding capacity regions were inspected.

APPENDIX A

PROOF OF THEOREM 2

A. Achievability

The direct proof is based on a coding scheme that achieves the corner points of the \mathcal{R}_{WAK} . The corner points are stated in (19a)-(19b) and illustrated in Fig 3. Fix a PMF $P_{X_1, X_2, Y} \in \mathcal{P}$, $\epsilon > 0$ and a PMF $P_{X_1, X_2} P_{V|X_1} P_{U|X_2, V} P_{Y|X_1, U, V}$ that has $P_{X_1, X_2, Y}$ as a marginal. Recall that $P_{X_1, X_2, Y}$ factors as $P_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1=f(Y)\}}$ and that it has the input PMF P_{X_1, X_2} as a marginal.

Codebook Generation: A codebook \mathcal{C}_V that comprises 2^{nR_V} codewords $\mathbf{v}(i)$, where $i \in [1 : 2^{nR_V}]$, each generated according to P_V^n . The codebook \mathcal{C}_V is randomly partitioned into $2^{nR_{12}}$ bins indexed by $t_{12} \in [1 : 2^{nR_{12}}]$ and denoted by $\mathcal{B}_V(t_{12})$. For every $\mathbf{v} \in \mathcal{C}_V$ a codebook $\mathcal{C}_U(\mathbf{v})$ is generated. Each codebook $\mathcal{C}_U(\mathbf{v})$ is assembled of 2^{nR_U} codewords $\mathbf{u}(\mathbf{v}, j)$, $j \in [1 : 2^{nR_U}]$, generated according to $P_{U|V}^n(\cdot|\mathbf{v})$. Each $\mathcal{C}_U(\mathbf{v})$ codebook is randomly partitioned into $2^{nR'_2}$ bins $\mathcal{B}_U(\mathbf{v}, t'_2)$, where $t'_2 \in [1 : 2^{nR'_2}]$. Moreover, the set $\mathcal{T}_\epsilon^{(n)}(P_{X_1})$ is partitioned into $2^{nR'_1}$ bins $\mathcal{B}_{X_1}(t'_1)$, where $t'_1 \in [1 : 2^{nR'_1}]$. To achieve (19a), consider the following scheme:

Encoding at Encoder 1: Upon receiving \mathbf{x}_1 , Encoder 1 searches a pair of indices $(i, t'_1) \in [1 : 2^{nR_V}] \times [1 : 2^{nR'_1}]$ such that $(\mathbf{v}(i), \mathbf{x}_1) \in \mathcal{T}_\epsilon^{(n)}(P_{V, X_1})$ and $\mathbf{x}_1 \in \mathcal{B}_{X_1}(t'_1)$. A concatenation i and t'_1 is conveyed to the decoder. The bin index of $\mathbf{v}(i)$, i.e., the index $t_{12} \in [1 : 2^{nR_{12}}]$ such that $\mathbf{v}(i) \in \mathcal{B}_V(t_{12})$, is conveyed to Encoder 2 via the cooperation link. Having

$$R_V > I(V; X_1) \quad (32)$$

ensures that such a codeword $\mathbf{v}(i)$ is found with high probability.

Decoding at Encoder 2: Given the source sequence \mathbf{x}_2 and the bin index t_{12} , Encoder 2 searches for an index $\hat{i} \in [1 : 2^{nR_V}]$ such that $\mathbf{v}(\hat{i}) \in \mathcal{B}_V(t_{12})$ and $(\mathbf{v}(\hat{i}), \mathbf{x}_2) \in \mathcal{T}_\epsilon^{(n)}(P_{V, X_2})$. Reliable decoding follows by having

$$R_V - R_{12} < I(V; X_2). \quad (33)$$

Encoding at Encoder 2: After decoding $\mathbf{v}(\hat{i})$, Encoder 2 searches for an index $j \in [1 : 2^{nR_U}]$, such that $\mathbf{u}(\mathbf{v}(\hat{i}), j) \in \mathcal{C}_U(\mathbf{v}(\hat{i}))$ and $(\mathbf{v}(\hat{i}), \mathbf{u}(\mathbf{v}(\hat{i}), j), \mathbf{x}_2) \in \mathcal{T}_\epsilon^{(n)}(P_{V, U, X_2})$. The bin number of the chosen $\mathbf{u}(\mathbf{v}(\hat{i}), j)$, that is, the index $t'_2 \in [1 : 2^{nR'_2}]$ such that $\mathbf{u}(\mathbf{v}(\hat{i}), j) \in \mathcal{B}_U(\mathbf{v}(\hat{i}), t'_2)$, is conveyed to the decoder. Choosing

$$R_U > I(U; X_2|V) \quad (34)$$

ensures that such a codeword $\mathbf{u}(\mathbf{v}(\hat{i}), j)$ is found with high probability.

Decoding and Output Generation: Upon receiving (i, t'_1) from Encoder 1 and t'_2 from Encoder 2, the decoder first identifies the codeword $\mathbf{v}(i) \in \mathcal{C}_V$ that is associated with i . Then it searches the bin $\mathcal{B}_{X_1}(t'_1)$ for a sequence $\hat{\mathbf{x}}_1$ such that $(\mathbf{v}(i), \hat{\mathbf{x}}_1) \in \mathcal{T}_\epsilon^{(n)}(P_{V, X_1})$. Reliable lossless reconstruction of \mathbf{x}_1 follows, provided that

$$R'_1 > H(X_1|V). \quad (35)$$

Given $(\mathbf{v}(i), \hat{\mathbf{x}}_1)$, the decoder searches for an index $\hat{j} \in [1 : 2^{nR_u}]$, such that $\mathbf{u}(\mathbf{v}(i), \hat{j}) \in \mathcal{B}_U(\mathbf{v}(i), t'_2)$ and $(\mathbf{v}(i), \mathbf{u}(\mathbf{v}(i), \hat{j}), \hat{\mathbf{x}}_1) \in \mathcal{T}_\epsilon^{(n)}(P_{V, U, X_1})$. To insure error-free decoding, we take

$$R_U - R'_2 < I(U; X_1|V). \quad (36)$$

Finally, an output sequence \mathbf{y} is generated according to $P_{Y|X_1, U, V}^n(\cdot | \hat{\mathbf{x}}_1, \mathbf{u}(\mathbf{v}(i), \hat{j}), \mathbf{v}(i))$. By the choice of the PMF, this output sequence admits the desired coordination constraint.

By taking $(R_1, R_2) = (R'_1 + R_V, R'_2)$ and applying the Fourier-Motzkin elimination (FME) on (32)-(36), we obtain the rate bounds

$$R_{12} > I(V; X_1) - I(V; X_2) = I(V; X_1|X_2) \quad (37a)$$

$$R_1 > H(X_1|V) + I(V; X_1) = H(X_1) \quad (37b)$$

$$R_2 > I(U; X_2|V) - I(U; X_1|V) = I(U; X_2|X_1, V) \quad (37c)$$

which imply that (19a) is achievable.

To establish the achievability of (19b), the codebooks $\mathcal{C}_U(\mathbf{v})$, where $\mathbf{v} \in \mathcal{C}_V$, do not need binning.

Encoding at Encoder 1: Given \mathbf{x}_1 , Encoder 1 finds $\mathbf{v}(i) \in \mathcal{C}_V$ in a similar manner and conveys its bin index t_{12} to Encoder 2. Moreover, it conveys the bin index of the received \mathbf{x}_1 , say t'_1 , to the decoder. Again, by having (32), such a codeword $\mathbf{v}(i)$ is found with high probability.

Decoding at Encoder 2: Performed in a similar manner as before. We again take (33) to ensure reliable decoding.

Encoding at Encoder 2: Encoder 2 finds a codeword $\mathbf{u}(\mathbf{v}(\hat{i}), j) \in \mathcal{C}_U(\mathbf{v}(\hat{i}))$ in a manner similar to that presented in the previous scheme. Now, however, it sends to the decoder a concatenation of \hat{i} and j . This decoding process has a vanishing probability of error if (34) holds.

Decoding and Output Generation: Upon receiving t'_1 and (\hat{i}, j) from Encoder 1 and 2, respectively, the decoder first finds the $\mathbf{v}(\hat{i}) \in \mathcal{C}_V$ that is associated with \hat{i} and the $\mathbf{u}(\mathbf{v}(\hat{i}), j) \in \mathcal{C}_U(\mathbf{v}(\hat{i}))$ that is associated with $(\mathbf{v}(\hat{i}), j)$. Given $(\mathbf{v}(\hat{i}), \mathbf{u}(\mathbf{v}(\hat{i}), j))$, it searches the bin $\mathcal{B}_{X_1}(t'_1)$ for a sequence $\hat{\mathbf{x}}_1$ such that $(\mathbf{v}(\hat{i}), \mathbf{u}(\mathbf{v}(\hat{i}), j), \hat{\mathbf{x}}_1) \in \mathcal{T}_\epsilon^{(n)}(P_{V, U, X_1})$. Reliable lossless reconstruction of \mathbf{x}_1 is ensured provided

$$R'_1 > H(X_1|V, U). \quad (38)$$

Finally, an output sequence \mathbf{y} is generated in the same manner as in the coding scheme for (19a).

Taking $(R_1, R_2) = (R'_1, R_V + R_U)$ and applying the FME procedure on (32)-(34) and (38) yields the following

bounds:

$$R_{12} > I(V; X_1) - I(V; X_2) = I(V; X_1 | X_2) \quad (39a)$$

$$R_1 > H(X_1 | V, U) \quad (39b)$$

$$R_2 > I(V; X_1) + I(U; X_2 | V). \quad (39c)$$

This concludes the proof of achievability for (19b).

B. Converse

We show that given an achievable rate triple (R_{12}, R_1, R_2) , there exists a PMF $P_{X_1, X_2} P_{V|X_1} P_{U|X_2, V} P_{Y|X_1, U, V}$ that has $P_{X_2} P_{Y|X_2} \mathbb{1}_{\{X_1=f(Y)\}}$ as a marginal, such that the inequalities in (13) are satisfied. Since (R_{12}, R_1, R_2) is achievable, X_1^n can be reconstructed at the decoder with a small probability of error. By Fano's inequality we have

$$H(X_1^n | T_1, T_2) \leq n\epsilon_n, \quad (40)$$

Moreover, by the structure of the PMF, we rewrite the mutual information measure in (10c) as

$$R_2 \geq I(U; X_2 | X_1, V) \stackrel{(a)}{=} I(V; X_2 | X_1) + I(U; X_2 | X_1, V) = I(V, U; X_2 | X_1). \quad (41)$$

where (a) follows from the Markov relation $V - X_1 - X_2$.

For the lower bound on R_{12} , consider

$$\begin{aligned} nR_{12} &\geq H(T_{12}) \\ &\stackrel{(a)}{=} I(T_{12}; X_1^n, X_2^n) \\ &= \sum_{i=1}^n I(T_{12}; X_1^n, X_{2,i} | X_{2,i+1}^n) \\ &\geq \sum_{i=1}^n I(T_{12}; X_{1,i}, X_{2,i} | X_1^{n \setminus i}, X_{2,i+1}^n) \\ &\stackrel{(b)}{=} \sum_{i=1}^n I(T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n; X_{1,i}, X_{2,i}) \\ &\geq \sum_{i=1}^n I(T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n; X_{1,i} | X_{2,i}) \\ &\stackrel{(c)}{=} \sum_{i=1}^n I(V_i; X_{1,i} | X_{2,i}) \end{aligned} \quad (42)$$

where:

- (a) follows because T_{12} is a function of X_1^n ;
- (b) follows because the (X_1^n, X_2^n) are pairwise i.i.d.;
- (c) follows by defining $V_i \triangleq (T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n)$.

Next, for R_1 we have

$$\begin{aligned} nR_1 &\geq H(T_1) \\ &\geq H(T_1|T_2) \end{aligned} \tag{43}$$

$$\begin{aligned} &\stackrel{(a)}{=} I(T_1; X_1^n|T_2) \\ &= H(X_1^n|T_2) - H(X_1^n|T_1, T_2) \end{aligned}$$

$$\stackrel{(b)}{\geq} \sum_{i=1}^n H(X_{1,i}|T_2, X_1^{i-1}) - n\epsilon_n \tag{44}$$

$$\begin{aligned} &\geq \sum_{i=1}^n H(X_{1,i}|T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n) - n\epsilon_n \\ &\stackrel{(c)}{=} \sum_{i=1}^n H(X_{1,i}|V_i, U_i) - n\epsilon_n \end{aligned} \tag{45}$$

where:

- (a) follows because T_1 is a function of X_1^n ;
- (b) follows from (40) and the mutual information chain rule;
- (c) follows by defining $U_i \triangleq T_2$ and from the definition of V_i .

The rate bound on R_2 follows from

$$\begin{aligned} nR_2 &\geq H(T_2) \\ &\geq H(T_2|X_1^n) \\ &\geq I(T_2; X_2^n|X_1^n) \\ &= \sum_{i=1}^n I(T_2; X_{2,i}|X_1^{n \setminus i}, X_{2,i+1}^n, X_{1,i}) \\ &\stackrel{(a)}{=} \sum_{i=1}^n I(T_2, X_1^{n \setminus i}, X_{2,i+1}^n; X_{2,i}|X_{1,i}) \\ &\stackrel{(b)}{=} \sum_{i=1}^n I(T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n; X_{2,i}|X_{1,i}) \\ &\stackrel{(c)}{=} \sum_{i=1}^n I(V_i, U_i; X_{2,i}|X_{1,i}) \end{aligned} \tag{46}$$

where:

- (a) follows because the (X_1^n, X_2^n) are pairwise i.i.d.;
- (b) follows because T_{12} is a function of X_1^n ;
- (c) follows since conditioning cannot increase entropy and from the definition of (V_i, U_i) .

For the sum of rates, we begin by writing

$$n(R_1 + R_2) \geq H(T_1, T_2) = H(T_2) + H(T_1|T_2). \tag{47}$$

For the first term in (47) we have

$$\begin{aligned}
H(T_2) &\geq I(T_2; X_1^n, X_2^n) \\
&= \sum_{i=1}^n I(T_2; X_1^n, X_{2,i} | X_{2,i+1}^n) \\
&\geq \sum_{i=1}^n I(T_2; X_{1,i}, X_{2,i} | X_1^{n \setminus i}, X_{2,i+1}^n) \\
&\stackrel{(a)}{=} \sum_{i=1}^n I(T_2, X_1^{n \setminus i}, X_{2,i+1}^n; X_{1,i}, X_{2,i}) \\
&= \sum_{i=1}^n [I(T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n; X_{1,i}, X_{2,i}) - I(T_{12}; X_{1,i}, X_{2,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n)] \\
&\stackrel{(b)}{=} \sum_{i=1}^n [I(V_i, U_i; X_{1,i}, X_{2,i}) - H(X_{1,i}, X_{2,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n) + H(X_{1,i}, X_{2,i} | T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n)] \\
&\stackrel{(c)}{=} \sum_{i=1}^n [I(V_i, U_i; X_{1,i}, X_{2,i}) + H(X_{1,i} | V_i, U_i) - H(X_{1,i}, X_{2,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n) \\
&\quad + H(X_{2,i} | T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n, X_{1,i})] \\
&\stackrel{(d)}{=} \sum_{i=1}^n [I(V_i, U_i; X_{1,i}, X_{2,i}) + H(X_{1,i} | V_i, U_i) - H(X_{1,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n) \\
&\quad - H(X_{2,i} | T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n, X_{1,i}) + H(X_{2,i} | T_2, T_{12}, X_1^{n \setminus i}, X_{2,i+1}^n, X_{1,i})] \\
&\stackrel{(e)}{=} \sum_{i=1}^n [I(V_i, U_i; X_{1,i}, X_{2,i}) + H(X_{1,i} | V_i, U_i) - \Delta_i] \tag{48}
\end{aligned}$$

where:

(a) follows because the (X_1^n, X_2^n) are pairwise i.i.d.;

(b) and (c) follow from the definition of (V_i, U_i) ;

(d) follows because T_{12} is a function of X_1^n ;

(e) follows by defining

$$\Delta_i \triangleq H(X_{1,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n), \tag{49}$$

The second term in (47) is lower bounded as

$$\begin{aligned}
H(T_1 | T_2) &\stackrel{(a)}{\geq} \sum_{i=1}^n H(X_{1,i} | T_2, X_1^{i-1}) - n\epsilon_n \\
&\geq \sum_{i=1}^n H(X_{1,i} | T_2, X_1^{n \setminus i}, X_{2,i+1}^n) - \epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \Delta_i \tag{50}
\end{aligned}$$

where (a) follows by repeating steps (43)-(44) in the lower bounding of R_1 , while (b) follows from (49).

By inserting (48) and (50) into (47), we have

$$n(R_1 + R_2) \geq \sum_{i=1}^n \left[I(V_i, U_i; X_{1,i}, X_{2,i}) + H(X_{1,i} | V_i, U_i) \right] - n\epsilon_n. \quad (51)$$

The upper bounds in (42), (45), (46) and (51) are rewritten by introducing a time-sharing random variable Q that is uniformly distributed over the set $[1 : n]$. The rate bound on R_{12} is rewritten as

$$R_{12} \geq \frac{1}{n} \sum_{q=1}^n I(V_q; X_{1,q} | X_{2,q}, Q = q) \quad (52)$$

$$= \sum_{q=1}^n \mathbb{P}(Q = q) I(V_q; X_{1,q} | X_{2,q}, Q = q) \quad (53)$$

$$= I(V_Q; X_{1,Q} | X_{2,Q}, Q) \quad (54)$$

$$\stackrel{(a)}{=} I(V_Q, Q; X_{1,Q} | X_{2,Q}) \quad (55)$$

where (a) follows because Q is independent of the pair $(X_{1,Q}, X_{2,Q})$ (see property 1 in [13, Section IIV-B]). By rewriting (45), (46) and (51) in the same manner, the region obtained is convex. This follows from the presence of the time-sharing random variable Q in the conditioning of all the mutual information and entropy terms.

Next, let $V \triangleq (V_Q, Q)$, $U \triangleq U_Q$. By using the time-mixing properties (property 2 in [13, Section IIV-B]), we have

$$R_{12} \geq I(V; X_1 | X_2) \quad (56a)$$

$$R_1 \geq H(X_1 | V, U) - \epsilon_n \quad (56b)$$

$$R_2 \geq I(U; X_2 | X_1, V) \quad (56c)$$

$$R_1 + R_2 \geq H(X_1 | V, U) + I(V, U; X_1, X_2) - \epsilon_n. \quad (56d)$$

To complete the converse, the following Markov relations must be shown to hold.

$$V - X_1 - X_2 \quad (57a)$$

$$U - (X_2, V) - X_1 \quad (57b)$$

$$Y - (X_1, U, V) - X_2. \quad (57c)$$

We prove that the Markov relations in (57a)-(57c) hold for every $q \in [1 : n]$. Upon doing so, showing that the relations also hold in their single-letter (as stated in (57a)-(57c)) is straightforward.

For (57a), recall that $V_q = (T_{12}, X_1^{n \setminus q}, X_{2,q+1}^n)$, and consider

$$\begin{aligned} 0 &\leq I(T_{12}, X_1^{n \setminus q}, X_{2,q+1}^n; X_{2,q} | X_{1,q}) \\ &\stackrel{(a)}{=} I(X_1^{n \setminus q}, X_{2,q+1}^n; X_{2,q} | X_{1,q}) \stackrel{(b)}{=} 0 \end{aligned}$$

where (a) follows because T_{12} is a function of X_1^n , while (b) follows because (X_1^n, X_2^n) are pairwise i.i.d. Thus

(57a) holds.

To establish (57b), we use Lemma 1 in [56]. Since $U_q = T_2$ we have

$$\begin{aligned} 0 &\leq I(T_2; X_{1,q} | X_{2,q}, T_{12}, X_1^{n \setminus q}, X_{2,q+1}^n) \\ &\leq I(T_2; X_{1,q}, X_{1,q+1}^n | X_{2,q}, T_{12}, X_1^{q-1}, X_{2,q+1}^n). \end{aligned} \quad (58)$$

Set

$$\begin{aligned} A_1 &= X_1^{q-1}, \quad A_2 = (X_{1,q}, X_{1,q+1}^n), \\ B_1 &= X_2^{q-1}, \quad B_2 = (X_{2,q}, X_{2,q+1}^n). \end{aligned}$$

Accordingly, (58) is rewritten as

$$0 \leq I(T_2; A_2 | T_{12}, A_1, B_2). \quad (59)$$

By noting that $T_{12} = T_{12}(A_1, A_2)$, $T_2 = (T_{12}, B_1, B_2)$ and $P_{A_1, A_2, B_1, B_2} = P_{A_1, B_1} P_{A_2, B_2}$ and applying the result of [56, Lemma 1, Conclusion 2], we have

$$0 \leq I(T_2; A_2 | T_{12}, A_1, B_2) = 0. \quad (60)$$

This implies (57b).

Finally, to prove (57c), recall that Y^n (and, in particular, Y_q) is a function of (T_1, T_2) . Rewriting (57c) explicitly, we have

$$Y_q - (T_2, T_{12}, X_1^n, X_{2,q+1}^n) - X_{2,q}. \quad (61)$$

Since T_1 is defined by X_1^n , we add it to the middle link of the Markov chain in (61) and establish the desired Markov property. This concludes the converse.

APPENDIX B

PROOF OF THEOREM 6

A. Achievability

To establish the achievability of the region in (13), we show that for a fixed $\epsilon > 0$, a fixed PMF

$$P_{V,U,Y_1} P_{X|V,U,Y_1} P_{Y_2|X} \quad (62)$$

for which $Y_1 = f(X)$ and rates (R_{12}, R_1, R_2) that satisfy (13), there is a sufficiently large $n \in \mathbb{N}$ and a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code such that $P_e^{(n)} \leq \epsilon$. We first derive an achievable region for a general BC with a one-sided conferencing link between the decoders that is described using three auxiliaries (rather than two). Then, by a proper choice of the auxiliaries, we show that the obtained region reduces to (13). Fix a PMF

$$P_{V,U_1,U_2} P_{X|V,U_1,U_2} P_{Y_2|X} \quad (63)$$

for which $Y_1 = f(X)$ and $\epsilon > 0$, and consider the following coding scheme.

Codebook Generation: Split each message m_j , $j = 1, 2$, into two sub-messages denoted by (m_{j0}, m_{jj}) . The pair (m_{10}, m_{20}) is referred to as a *public message* while m_{jj} serve as *private message* j . The rates associated with m_{j0} and m_{jj} , $j = 1, 2$, are denoted by R_{j0} and R_{jj} , while the corresponding alphabets are \mathcal{M}_{j0} and \mathcal{M}_{jj} , respectively. Accordingly, we have

$$R_j = R_{j0} + R_{jj}, \quad j = 1, 2. \quad (64)$$

The random variables M_{j0} and M_{jj} are associated with the public part of message j and private message j , respectively. M_{j0} and M_{jj} are independent and uniform over \mathcal{M}_{j0} and \mathcal{M}_{jj} .

Generate a public message codebook, denoted by \mathcal{C}_V , that comprises $2^{n(R_{10}+R_{20})}$ v -codewords $\mathbf{v}(m_{10}, m_{20})$, $(m_{10}, m_{20}) \in \mathcal{M}_{10} \times \mathcal{M}_{20}$, each drawn according to P_V^n independent of all the other v -codewords. Randomly and uniformly partition the codebook \mathcal{C}_V into $2^{nR_{12}}$ equal-sized bins $\mathcal{B}(m_{12})$, where $m_{12} \in \mathcal{M}_{12}$.

For each $\mathbf{v}(m_{10}, m_{20}) \in \mathcal{C}_V$, generate two codebooks $\mathcal{C}_{U_j}(m_{10}, m_{20})$, $j = 1, 2$, each comprises $2^{n(R_{jj}+R'_j)}$ codewords \mathbf{u}_j , each drawn according to $P_{U_j|V}^n(\cdot | \mathbf{v}(m_{10}, m_{20}))$ independent of all the other u_j -codewords. The u_j -codewords in $\mathcal{C}_{U_j}(m_{10}, m_{20})$ are labeled as $\mathbf{u}_j(m_{10}, m_{20}, m_{jj}, i_j)$, where $(m_{jj}, i_j) \in \mathcal{M}_{jj} \times \mathcal{I}_j$ and $\mathcal{I}_j = [1 : 2^{nR'_j}]$. Based on this labeling, the codebook $\mathcal{C}_{U_j}(m_{10}, m_{20})$ has a u_j -bin associated with every $m_{jj} \in \mathcal{M}_{jj}$, each containing $2^{nR'_j}$ u_j -codewords.

Encoding: To transmit the message pair $(m_1, m_2) = ((m_{10}, m_{11}), (m_{20}, m_{22}))$, the encoder searches for a pair $(i_1, i_2) \in \mathcal{I}_1 \times \mathcal{I}_2$, such that

$$\left(\mathbf{v}(m_{10}, m_{20}), \mathbf{u}_1(m_{10}, m_{20}, m_{11}, i_1), \mathbf{u}_2(m_{10}, m_{20}, m_{22}, i_2) \right) \in \mathcal{T}_\epsilon^{(n)}(P_{V,U_1,U_2}) \quad (65)$$

where $\mathbf{v}(m_{10}, m_{20}) \in \mathcal{C}_V$ and $\mathbf{u}_j(m_{10}, m_{20}, m_{jj}, i_j) \in \mathcal{C}_{U_j}(m_{10}, m_{20})$, for $j = 1, 2$. If the set of appropriate index pairs contains more than one element, the encoder chooses a pair at random according to a uniform distribution over that set; if the set is empty, a pair is chosen uniformly over $\mathcal{I}_1 \times \mathcal{I}_2$. The channel input sequence \mathbf{x} is then randomly generated according to $P_{X|V,U_1,U_2}^n$ and is transmitted over the channel.

Decoding Process: *Decoder 1:* Searches for a unique triple $(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{11}) \in \mathcal{M}_{10} \times \mathcal{M}_{20} \times \mathcal{M}_{11}$ for which there is an index $\hat{i}_1 \in \mathcal{I}_1$, such that

$$\left(\mathbf{v}(\hat{m}_{10}, \hat{m}_{20}), \mathbf{u}_1(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{11}, \hat{i}_1), \mathbf{y}_1 \right) \in \mathcal{T}_\epsilon^{(n)}(P_{V,U_1,Y_1}) \quad (66)$$

where $\mathbf{v}(\hat{m}_{10}, \hat{m}_{20}) \in \mathcal{C}_V$ and $\mathbf{u}_1(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{11}, \hat{i}_1) \in \mathcal{C}_{U_1}(\hat{m}_{10}, \hat{m}_{20})$. If such a unique triple is found, then $\hat{m}_1 = (\hat{m}_{10}, \hat{m}_{11})$ is declared as the decoded message; otherwise, an error is declared.

Cooperation: Given $(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{11}, \hat{i}_1)$, Decoder 1 conveys the bin number of $\mathbf{v}(\hat{m}_{10}, \hat{m}_{20})$ to Decoder 2 via the cooperation link. Namely, Decoder 1 shares with Decoder 2 the index $\hat{m}_{12} \in \mathcal{M}_{12}$, such that $\mathbf{v}(\hat{m}_{10}, \hat{m}_{20}) \in \mathcal{B}(\hat{m}_{12})$.

Decoder 2: Upon receiving \hat{m}_{12} from Decoder 1 and \mathbf{y}_2 from the channel, Decoder 2 searches for a unique triple

$(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{22}) \in \mathcal{M}_{10} \times \mathcal{M}_{20} \times \mathcal{M}_{22}$ for which there is an $\hat{i}_2 \in \mathcal{I}_2$, such that

$$\left(\mathbf{v}(\hat{m}_{10}, \hat{m}_{20}), \mathbf{u}_2(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{22}, \hat{i}_2), \mathbf{y}_2 \right) \in \mathcal{T}_\epsilon^{(n)}(P_{V,U_2,Y_2}) \quad (67)$$

where $\mathbf{v}(\hat{m}_{10}, \hat{m}_{20}) \in \mathcal{B}(\hat{m}_{12})$ and $\mathbf{u}_2(\hat{m}_{10}, \hat{m}_{20}, \hat{m}_{22}, \hat{i}_2) \in \mathcal{C}_{U_2}(\hat{m}_{10}, \hat{m}_{20})$. If such a unique triple is found, then $\hat{m}_2 \triangleq (\hat{m}_{20}, \hat{m}_{22})$ is declared as the decoded message; otherwise, an error is declared.

By standard error probability analysis (see Appendix C), reliability is achieved provided that

$$R'_1 + R'_2 > I(U_1; U_2|V) \quad (68a)$$

$$R_{11} + R'_1 < I(U_1; Y_1|V) \quad (68b)$$

$$R_{20} + R_1 + R'_1 < I(V, U_1; Y_1) \quad (68c)$$

$$R_{22} + R'_2 < I(U_2; Y_2|V) \quad (68d)$$

$$R_{10} + R_2 + R'_2 - R_{12} < I(V, U_2; Y_2). \quad (68e)$$

Applying FME on (68) while using (64) yields the rate bounds

$$R_1 < I(V, U_1; Y_1) \quad (69a)$$

$$R_2 < I(V, U_2; Y_2) + R_{12} \quad (69b)$$

$$R_1 + R_2 < I(V, U_1; Y_1) + I(U_2; Y_2|V) - I(U_1; U_2|V) \quad (69c)$$

$$R_1 + R_2 < I(U_1; Y_1|V) + I(V, U_2; Y_2) - I(U_1; U_2|V) + R_{12}. \quad (69d)$$

By setting $U_1 = Y_1$ and relabeling U_2 as U , the bounds in (69) reduce to (13). Note that this choice of auxiliaries is valid as they satisfy the Markov relations stated in Theorem 6. This shows that \mathcal{C}_{BC} is achievable.

Remark 16 *The cooperation protocol described in the proof is reminiscent of the WZ coding technique. The cooperation link is used to convey a bin of the common message codeword \mathbf{v} (rather than the codeword itself) from 1st decoder to the 2nd. As part of the joint typicality decoding rule in (67), the channel input \mathbf{y}_2 is used as correlated side information to isolate the actual v -codeword from the bin. The correlation between the random sequences \mathbf{V} and \mathbf{Y}_2 is induced from the channel transition probability and the underlying Markov relations (with respect to the PMF in (63)).*

B. Converse

We show that if a rate triple (R_{12}, R_1, R_2) is achievable, then there exists a PMF $P_{V,U,Y_1,X}P_{Y_2|X}$ for which $Y_1 = f(X)$, such that the inequalities in (13) are satisfied. Since (R_{12}, R_1, R_2) is achievable, for every $\epsilon > 0$, there is a sufficiently large n and a $(n, 2^{nR_{12}}, 2^{nR_1}, 2^{nR_2})$ code for which $P_e^{(n)} \leq \epsilon$. By Fano's inequality we have

$$H(M_1|Y_1^n) \leq nR_1P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n^{(1)}, \quad (70a)$$

$$H(M_2|M_{12}, Y_2^n) \leq nR_2 P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n^{(2)}, \quad (70b)$$

where $\lim_{n \rightarrow \infty} \epsilon_n^{(j)} = 0$, for $j = 1, 2$. Define

$$\epsilon_n = \max \{\epsilon_n^{(1)}, \epsilon_n^{(2)}\}. \quad (71)$$

It follows that

$$\begin{aligned} nR_1 &= H(M_1) \\ &\stackrel{(a)}{\leq} I(M_1; Y_1^n) + n\epsilon_n \\ &\stackrel{(b)}{=} I(X^n; Y_1^n) + n\epsilon_n \\ &\stackrel{(c)}{\leq} \sum_{i=1}^n H(Y_{1,i}) + n\epsilon_n \end{aligned} \quad (72)$$

where:

(a) follows from (70a) and (71);

(b) follows since $M_1 - X^n - Y_1^n$ forms a Markov chain and by the Data Processing Inequality;

(c) follows because Y_1^n is a function of X^n , by the mutual information chain rule and since conditioning cannot increase entropy.

To bound R_2 consider

$$nR_2 = H(M_2) \quad (73)$$

$$\begin{aligned} &\stackrel{(a)}{\leq} I(M_2; M_{12}, Y_2^n) + n\epsilon_n \\ &= I(M_2; Y_2^n | M_{12}) + I(M_2; M_{12}) + n\epsilon_n \end{aligned} \quad (74)$$

$$\begin{aligned} &\leq \sum_{i=1}^n I(M_2; Y_{2,i} | M_{12}, Y_{2,i+1}^n) + H(M_{12}) + n\epsilon_n \\ &\stackrel{(b)}{\leq} \sum_{i=1}^n I(V_i, U_i; Y_{2,i}) + nR_{12} + n\epsilon_n \end{aligned} \quad (75)$$

where (a) follows from (70b) and (71), while (b) follows by defining $V_i \triangleq (M_{12}, Y_1^{i-1}, Y_{2,i+1}^n)$ and $U_i \triangleq M_2$ and because a uniform distribution maximizes entropy.

For the sum of rates, we begin by writing

$$n(R_1 + R_2) = H(M_1, M_2) = H(M_2) + H(M_1 | M_2). \quad (76)$$

By the independence of M_1 and M_2 and by (70a) and (71), we have

$$H(M_1 | M_2) \leq H(Y_1^n | M_2) + n\epsilon_n. \quad (77)$$

Moreover, we can bound

$$\begin{aligned}
H(M_2) &\stackrel{(a)}{\leq} I(M_2; Y_2^n | M_{12}) + I(M_2; M_{12}) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \left[I(M_2; Y_{2,i}^n | M_{12}, Y_1^{i-1}) - I(M_2; Y_{2,i+1}^n | M_{12}, Y_1^i) \right] + I(M_2; M_{12}) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n \left[I(M_2; Y_{2,i+1}^n | M_{12}, Y_1^{i-1}) + I(U_i; Y_{2,i} | V_i) - I(M_2; Y_{1,i}, Y_{2,i+1}^n | M_{12}, Y_1^{i-1}) \right. \\
&\quad \left. + I(M_2; Y_{1,i} | M_{12}, Y_1^{i-1}) \right] + I(M_2; M_{12}) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n \left[I(U_i; Y_{2,i} | V_i) - I(M_2; Y_{1,i} | M_{12}, Y_1^{i-1}, Y_{2,i+1}^n) \right] + I(M_2; M_{12}, Y_1^n) + n\epsilon_n \\
&\stackrel{(e)}{=} \sum_{i=1}^n \left[I(U_i; Y_{2,i} | V_i) - I(U_i; Y_{1,i} | V_i) \right] + I(M_2; Y_1^n) + n\epsilon_n
\end{aligned} \tag{78}$$

where:

- (a) follows from repeating steps (73)-(74) in the upper bounding of R_2 ;
- (b) follows from a telescoping identity [34, Eq. (9) and (11)];
- (c) follows from the definition of (V_i, U_i) ;
- (d) follows for the mutual information chain rule;
- (e) follows from the definition of (V_i, U_i) (second term) and from the Markov relation $M_{12} - Y_1^n - M_2$ (third term).

Inserting (77) and (78) into (76) results in

$$n(R_1 + R_2) \leq \sum_{i=1}^n \left[I(U_i; Y_{2,i} | V_i) - I(U_i; Y_{1,i} | V_i) \right] + H(Y_1^n) + 2n\epsilon_n \tag{79}$$

$$\stackrel{(a)}{\leq} \sum_{i=1}^n \left[H(Y_{1,i} | V_i, U_i) + I(U_i; Y_{2,i} | V_i) + I(V_i; Y_{1,i}) \right] + 2n\epsilon_n \tag{80}$$

where (a) follows because conditioning cannot increase entropy.

Finally, note that

$$\begin{aligned}
H(Y_1^n) - \sum_{i=1}^n H(Y_{1,i} | V_i) &\stackrel{(a)}{=} \sum_{i=1}^n I(Y_{2,i+1}^n; Y_{1,i} | M_{12}, Y_1^{i-1}) + I(M_{12}; Y_1^n) \\
&\stackrel{(b)}{\leq} \sum_{i=1}^n I(Y_1^{i-1}; Y_{2,i} | M_{12}, Y_{2,i+1}^n) + H(M_{12}) \\
&\stackrel{(c)}{\leq} \sum_{i=1}^n I(V_i; Y_{2,i}) + nR_{12}
\end{aligned} \tag{81}$$

where:

- (a) follows by the mutual information chain rule and the definition of V_i ;
- (b) follows by the Csiszár sum identity and the non-negativity of entropy;
- (c) follows because conditioning cannot increase entropy while a uniform distribution maximizes it.

By plugging (81) into (79), we obtain

$$n(R_1 + R_2) \leq \sum_{i=1}^n \left[H(Y_{1,i}|V_i, U_i) + I(V_i, U_i; Y_{2,i}) \right] + nR_{12} + 2n\epsilon_n \quad (82)$$

The upper bounds in (72), (75), (80) and (82) can be rewritten by introducing a time-sharing random variable Q that is uniformly distributed over the set $[1 : n]$. For instance, the bound in (75) is rewritten as

$$\begin{aligned} R_2 &\leq \frac{1}{n} \sum_{q=1}^n I(V_q, U_q; Y_{2,q}) + R_{12} + \epsilon_n \\ &= \sum_{q=1}^n \mathbb{P}(Q = q) I(V_q, U_q; Y_{2,q}|Q = q) + R_{12} + \epsilon_n \\ &= I(V_Q, U_Q; Y_{2,Q}|Q) + R_{12} + \epsilon_n \\ &\leq I(Q, V_Q, U_Q; Y_{2,Q}) + R_{12} + \epsilon_n. \end{aligned} \quad (83)$$

By rewriting the rate bounds (72), (80) and (82) in the same manner, the region obtained is convex. Next, let $Y_1 \triangleq Y_{1,Q}$, $Y_2 \triangleq Y_{2,Q}$, $V \triangleq (Q, V_Q)$ and $U \triangleq U_Q$. We have

$$R_1 \leq H(Y_1) + \epsilon_n \quad (84a)$$

$$R_2 \leq I(V, U; Y_2) + R_{12} + \epsilon_n \quad (84b)$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) + 2\epsilon_n \quad (84c)$$

$$R_1 + R_2 \leq H(Y_1|V, U) + I(V, U; Y_2) + R_{12} + 2\epsilon_n. \quad (84d)$$

To complete the proof we need to show that the PMF of (V, U, X, Y_1, Y_2) factors as $P_{V,U,Y_1,X}P_{Y_2|X}$, which boils down to the Markov relation

$$(V, U, Y_1) - X - Y_2. \quad (85)$$

The proof of (85) is given in Appendix H and establishes the converse.

APPENDIX C

ERROR PROBABILITY ANALYSIS FOR THE ACHIEVABILITY OF THEOREM 6

Encoding errors: Denote the transmitted messages by $(m_1, m_2) = ((m_{10}, m_{11}), (m_{20}, m_{22}))$. For brevity, we denote $m_0 \triangleq (m_{10}, m_{20})$ and describe an encoding error as:

$$\mathcal{L} = \left\{ \forall (i_1, i_2) \in \mathcal{I}_1 \times \mathcal{I}_2, (\mathbf{V}(m_1), \mathbf{U}_1(m_0, m_{11}, i_1), \mathbf{U}_2(m_0, m_{22}, i_2)) \notin \mathcal{T}_\epsilon^{(n)}(P_{V,U_1,U_2}) \right\}. \quad (86)$$

Namely, an encoding error occurs if there is no pair of indices $(i_1, i_2) \in \mathcal{I}_1 \times \mathcal{I}_2$ that satisfies (65). By the Multivariate Covering Lemma [57, Lemma 8.2], $\mathbb{P}(\mathcal{L}) \rightarrow 0$ as $n \rightarrow \infty$ if we have

$$R'_1 + R'_2 > I(U_1; U_2|V). \quad (87)$$

Decoding errors: To account for decoding errors, define the following event

$$\mathcal{E}_j(m_0, m_{jj}, i_j) = \left\{ (\mathbf{V}(m_0), \mathbf{U}_j(m_0, m_{jj}, i_j), \mathbf{Y}_j) \in \mathcal{T}_\epsilon^{(n)}(P_{V, U_j, Y_j}) \right\} \quad (88)$$

for $j = 1, 2$. Let $(i_1, i_2) \in \mathcal{I}_1 \times \mathcal{I}_2$ denote the pair of indices that were originally chosen by the encoder and (I_1, I_2) denote the corresponding random variables. Define

$$\mathcal{K} = \left\{ (M_1, M_2, I_1, I_2) = (m_1, m_2, i_1, i_2) \right\}. \quad (89)$$

By the union bound, the error probability when averaged over the ensemble of codebooks is bounded as

$$\begin{aligned} P_e^{(n)} \leq & \mathbb{P}(\mathcal{L}) + (1 - \mathbb{P}(\mathcal{L})) \left(\sum_{j=1}^2 \left[\underbrace{\mathbb{P}(\mathcal{E}_j^c(m_0, m_{jj}, i_j) | \mathcal{K})}_{P_j^{[1]}} + \underbrace{\mathbb{P} \left(\bigcup_{\hat{i}_j, \tilde{m}_{jj} \neq m_{jj}} \mathcal{E}_j(m_0, \tilde{m}_{jj}, \hat{i}_j) | \mathcal{K} \right)}_{P_j^{[2]}} \right] \right. \\ & + \underbrace{\mathbb{P} \left(\bigcup_{\tilde{m}_0 \neq m_0} \mathcal{E}_1(\tilde{m}_0, m_{11}, i_1) | \mathcal{K} \right)}_{P_1^{[3]}} + \underbrace{\mathbb{P} \left(\bigcup_{\hat{i}_1, (\tilde{m}_0, \tilde{m}_{11}) \neq (m_0, m_{11})} \mathcal{E}_1(\tilde{m}_0, \tilde{m}_{11}, \hat{i}_1) | \mathcal{K} \right)}_{P_1^{[4]}} \\ & \left. + \underbrace{\mathbb{P} \left(\bigcup_{\substack{\tilde{m}_0 \neq m_0: \\ \mathbf{V}(\tilde{m}_0) \in \mathcal{B}(m_{12})}} \mathcal{E}_2(\tilde{m}_0, m_{22}, i_2) | \mathcal{K} \right)}_{P_2^{[3]}} + \underbrace{\mathbb{P} \left(\bigcup_{\substack{\hat{i}_2, (\tilde{m}_0, \tilde{m}_{22}) \neq (m_0, m_{22}) \\ \mathbf{V}(\tilde{m}_0) \in \mathcal{B}(m_{12})}} \mathcal{E}_2(\tilde{m}_0, \tilde{m}_{22}, \hat{i}_2) | \mathcal{K} \right)}_{P_2^{[4]}} \right). \end{aligned} \quad (90)$$

Note that $\{P_j^{[k]}\}_{k=1}^4$ correspond to decoding errors by Decoder j , where $j = 1, 2$. We proceed with the following steps:

1) $P_j^{[1]}$, for $j = 1, 2$, vanishes to 0 as $n \rightarrow \infty$ by the law of large numbers.

2) To upper bound $P_1^{[j]}$, $j = 1, 2$, consider:

$$\begin{aligned} P_j^{[2]} & \stackrel{(a)}{\leq} \sum_{(\tilde{m}_{jj}, \tilde{i}_j) \neq (m_{jj}, i_j)} 2^{-n(I(U_j; Y_j | V) - \delta_\epsilon)} \\ & \leq 2^{n(R_{jj} + R'_j)} 2^{-n(I(U_j; Y_j | V) - \delta_\epsilon)} \\ & = 2^{n(R_{jj} + R'_j - I(U_j; Y_j | V) + \delta_\epsilon)} \end{aligned}$$

where (a) follows since for every $(\tilde{m}_{jj}, \tilde{i}_j) \neq (m_{jj}, i_j)$, $\mathbf{U}_j(m_0, \tilde{m}_{jj}, \tilde{i}_j)$ is independent of \mathbf{Y}_j while both of them are drawn conditioned on $\mathbf{V}(m_0)$. Moreover, $\delta_\epsilon \rightarrow 0$ as $n \rightarrow \infty$. Hence, to ensure that $P_j^{[2]}$ vanishes as $n \rightarrow \infty$, we take:

$$R_{jj} + R'_j < I(U_j; Y_j | V), \quad j = 1, 2. \quad (91)$$

3) For $P_1^{[4]}$, we have:

$$\begin{aligned}
P_1^{[4]} &\stackrel{(a)}{\leq} \sum_{\substack{(\tilde{m}_0, \tilde{m}_{11}, \tilde{i}_1) \\ \neq (m_0, m_{11}, i_1)}} 2^{-n(I(V, U_1; Y_1) - \delta_\epsilon)} \\
&\leq 2^{n(R_{20} + R_1 + R'_1)} \cdot 2^{-n(I(V, U_1; Y_1) - \delta_\epsilon)} \\
&= 2^{n(R_{20} + R_1 + R'_1 - I(V, U_1; Y_1) + \delta_\epsilon)}
\end{aligned}$$

where (a) follows since for every $(\tilde{m}_0, \tilde{m}_{11}, \tilde{i}_1) \neq (m_0, m_{11}, i_1)$, $\mathbf{V}(\tilde{m}_0)$ and $\mathbf{U}_1(\tilde{m}_0, \tilde{m}_{11}, \tilde{i}_1)$ are drawn together by independent of \mathbf{Y}_1 . Again, $\delta_\epsilon \rightarrow 0$ as $n \rightarrow \infty$, and therefore, we have that $P_1^{[4]} \rightarrow 0$ as $n \rightarrow \infty$ if

$$R_{20} + R_1 + R'_1 < I(V, U_1; Y_1). \quad (92)$$

4) By repeating similar arguments as before and keeping in mind that the search space of m_0 at Decoder 2 is of size $2^{n(R_{10} + R_{20} - R_{12})}$ (due to the binning of \mathcal{C}_V and the cooperation protocol), we have that $P_2^{[4]}$ decays with n provided that

$$R_{10} + R_2 + R'_2 - R_{12} < I(V, U_2; Y_2). \quad (93)$$

5) By repeating similar steps to upper bound $P_1^{[3]}$, the obtained rate bound is redundant. This is since for every $\tilde{m}_0 \neq m_0$ the codewords $\mathbf{V}(\tilde{m}_0)$ and $\mathbf{U}_1(\tilde{m}_0, m_{11}, i_1)$ are independent of \mathbf{Y}_1 . Hence, to insure that $P_1^{[3]}$ vanishes to 0 as $n \rightarrow \infty$, we take

$$R_{10} + R_{20} < I(V, U_1; Y_1) \quad (94)$$

in which the right-hand side (RHS) coincides with the RHS of (92), while the left-hand side (LHS) is with respect to $R_{10} + R_{20}$ only. Clearly, (92) is the dominating constraint.

Summarizing the above results, we get that the RHS of (90) decays as the blocklength $n \rightarrow \infty$ if the conditions in (68) are met.

APPENDIX D

PROOF OF LEMMA 11

To show $\mathcal{C}_{BC}^{(D)} \subseteq \mathcal{C}_{BC}$, let $(R_{12}, R_1, R_2) \in \mathcal{C}_{BC}^{(D)}$ be a rate triple achieved by (V, U, X) . By taking $V^* = V$ and $U^* = U$, it follows that the same rate triple (R_{12}, R_1, R_2) is contained in \mathcal{C}_{BC} , as it is achieved by (V^*, U^*, X) (since substituting (17a) into (17d) yields (13d)).

Next, we establish that $\mathcal{C}_{BC} \subseteq \mathcal{C}_{BC}^{(D)}$. Let $(R_{12}, R_1, R_2) \in \mathcal{C}_{RBC}$ be a rate triple achieved by (V, U, X) . Further assume that

$$R_{12} < I(V; Y_1) - I(V; Y_2) \quad (95)$$

(otherwise, all four inequalities in (17) clearly hold). Accordingly, there is a real number $\gamma > 0$, such that

$$R_{12} = I(V; Y_1) - I(V; Y_2) - \gamma. \quad (96)$$

Define $V^* \triangleq (\Theta, \tilde{V})$, where $\Theta \sim \text{Ber}(\lambda)$, $\lambda \in [0, 1]$, is a binary random variable independent of (V, U, X) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$, and

$$\tilde{V} = \begin{cases} V, & \Theta = \theta_1 \\ \emptyset, & \Theta = \theta_2 \end{cases}. \quad (97)$$

Furthermore, let

$$\lambda = \frac{I(V; Y_1) - I(V; Y_2) - \gamma}{I(V; Y_1) - I(V; Y_2)} \quad (98)$$

and define $U^* = (V, U)$.

With respect to this choice of (V^*, U^*) , consider

$$I(V^*; Y_1) - I(V^*; Y_2) = \lambda [I(V; Y_1) - I(V; Y_2)] \stackrel{(a)}{=} I(V; Y_1) - I(V; Y_2) - \gamma \stackrel{(b)}{=} R_{12} \quad (99)$$

where (a) follows from the choice of λ in (98) and (b) follows from (96). Thus, (17a) is satisfied.

Next, by the definition of U^* and because (13a)-(13b) are valid, we obtain (17b)-(17c). Thus, it remains to show that (17d) holds. Consider the following steps:

$$\begin{aligned} H(Y_1|V^*, U^*) + I(U^*; Y_2|V^*) + I(V^*; Y_1) &= H(Y_1|V^*, U^*) + I(V^*, U^*; Y_2) + I(V^*; Y_1) - I(V^*; Y_2) \\ &\stackrel{(a)}{=} H(Y_1|V, U) + I(V, U; Y_2) + R_{12} \stackrel{(b)}{\geq} R_1 + R_2 \end{aligned} \quad (100)$$

where (a) follows from the definition of U^* and (99), while (b) follows from (13d). This implies that (17d) is also satisfied and yields the desired inclusion, that is, $\mathcal{C}_{BC} \subseteq \mathcal{C}_{BC}^{(D)}$.

APPENDIX E

EXPLICIT CONVERSE FOR LEMMA 11

The converse for Theorem 6 is established using a novel approach that generalizes the classical technique used for converse proofs. Our approach relies on two key properties. First, the construction of the auxiliary random variables depends on the distribution induced by the codebook. Second, the auxiliaries are constructed in a probabilistic manner.

We show that if a rate triple (R_{12}, R_1, R_2) is achievable, then there is a PMF $P_{V,U,Y_1,X} P_{Y_2|X}$ for which $Y_1 = f(X)$, such that the inequalities in (17) are satisfied. To do so, we first provide an upper bound on $\mathcal{C}_{BC}^{(D)}$ and then establish its inclusion in $\mathcal{C}_{BC}^{(D)}$. The upper bound is stated in the following lemma.

Lemma 17 (Upper Bound on the Capacity Region) *Let \mathcal{R}_O be the region defined by the union of rate triples*

$(R_{12}, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_{12} \geq I(A; Y_1|C) - I(C; Y_2|A) \quad (101a)$$

$$R_1 \leq H(Y_1|B, C) \quad (101b)$$

$$R_2 \leq I(B; Y_2|A) + R_{12} \quad (101c)$$

$$R_1 + R_2 \leq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C) \quad (101d)$$

where the union is over all PMFs $P_{A,B,C,Y_1,X}P_{Y_2|X}$ for which $Y_1 = f(X)$. The following inclusion holds:

$$\mathcal{C}_{BC}^{(D)} \subseteq \mathcal{R}_O. \quad (102)$$

Proof: By similar arguments to those given in Subsection B of Appendix B, since (R_{12}, R_1, R_2) is achievable and by Fano's inequality, we have

$$H(M_1|Y_1^n) \leq n\epsilon_n \quad (103a)$$

$$H(M_2|M_{12}, Y_2^n) \leq n\epsilon_n \quad (103b)$$

where $\lim_{n \rightarrow \infty} \epsilon_n = 0$. It follows that

$$\begin{aligned} nR_{12} &\geq H(M_{12}) \\ &\stackrel{(a)}{=} I(M_{12}; Y_1^n) \\ &\stackrel{(b)}{=} \sum_{i=1}^n \left[I(M_{12}, Y_{2,i+1}^n; Y_1^i) - I(M_{12}, Y_{2,i}^n; Y_1^{i-1}) \right] \\ &\stackrel{(c)}{=} \sum_{i=1}^n \left[I(M_{12}, Y_{2,i+1}^n; Y_{1,i} | Y_1^{i-1}) - I(Y_1^{i-1}; Y_{2,i} | M_{12}, Y_{2,i+1}^n) \right] \\ &\stackrel{(d)}{=} \sum_{i=1}^n \left[I(A_i; Y_{1,i} | C_i) - I(C_i; Y_{2,i} | A_i) \right] \end{aligned} \quad (104)$$

where:

- (a) follows because M_{12} is a function of Y_1^n ;
- (b) follows from a telescoping identity [34, Eq. (9) and (11)];
- (c) follows by the mutual information chain rule;
- (d) follows by defining $A_i \triangleq (M_{12}, Y_{2,i+1}^n)$ and $C_i \triangleq Y_1^{i-1}$.

For the upper bound on R_1 , consider

$$\begin{aligned} nR_1 &= H(M_1) \\ &= H(M_1|M_2) \\ &\stackrel{(a)}{\leq} I(M_1; Y_1^n|M_2) + n\epsilon_n \\ &\stackrel{(b)}{=} H(Y_1^n|M_2) - H(Y_1^n|M_1, M_2, X^n) + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&\stackrel{(c)}{=} \sum_{i=1}^n H(Y_{1,i}|M_2, Y_1^{i-1}) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n H(Y_{1,i}|B_i, C_i) + n\epsilon_n
\end{aligned} \tag{105}$$

where:

- (a) follows from (103a);
- (b) follows because X^n is a function of (M_1, M_2) ;
- (c) follows because Y_1^n is a function of X^n ;
- (d) follows by defining $B_i \triangleq M_2$ and from the definition of C_i .

To bound R_2 we have

$$\begin{aligned}
nR_2 &\stackrel{(a)}{\leq} I(M_2; Y_2^n | M_{12}) + I(M_2; M_{12}) + n\epsilon_n \\
&\leq \sum_{i=1}^n I(M_2; Y_{2,i} | M_{12}, Y_{2,i+1}^n) + H(M_{12}) + n\epsilon_n \\
&\stackrel{(b)}{\leq} \sum_{i=1}^n I(B_i; Y_{2,i} | A_i) + nR_{12} + n\epsilon_n
\end{aligned} \tag{106}$$

where (a) follows by repeating steps (73)-(74) in Appendix B, while (b) follows from the definition of (A_i, B_i) and because a uniform distribution maximizes entropy.

Finally, for the sum of rates, we begin from step (79) in Appendix B and note that the auxiliaries in Appendix B can be rewritten in terms of (A_i, B_i, C_i) as $V_i = (A_i, C_i)$ and $U_i = B_i$. We thus have

$$\begin{aligned}
n(R_1 + R_2) &\leq \sum_{i=1}^n \left[I(B_i; Y_{2,i} | A_i, C_i) - I(B_i; Y_{1,i} | A_i, C_i) \right] + H(Y_1^n) + 2n\epsilon_n \\
&\stackrel{(a)}{=} \sum_{i=1}^n \left[H(Y_{1,i} | A_i, B_i, C_i) + I(B_i; Y_{2,i} | A_i, C_i) + I(A_i; Y_{1,i} | C_i) \right] + 2n\epsilon_n
\end{aligned} \tag{107}$$

where (a) follows from the mutual information chain rule and the definition of (A_i, B_i, C_i) .

By standard time-sharing arguments, we rewrite the bounds in (104)-(107) as

$$R_{12} \geq I(A; Y_1 | C) - I(C; Y_2 | A) \tag{108a}$$

$$R_1 \leq H(Y_1 | B, C) + \epsilon_n \tag{108b}$$

$$R_2 \leq I(B; Y_2 | A) + R_{12} + \epsilon_n \tag{108c}$$

$$R_1 + R_2 \leq H(Y_1 | A, B, C) + I(B; Y_2 | A, C) + I(A; Y_1 | C) + 2\epsilon_n. \tag{108d}$$

By similar arguments to those presented in Appendix B, the Markov relations stated in Lemma 17 are established. ■

Using Lemma 17, the converse is established by the following lemma.

Lemma 18 (Tightness of the Upper Bound) *The following inclusion holds:*

$$\mathcal{R}_O \subseteq \mathcal{C}. \quad (109)$$

Proof: Let $(R_{12}, R_1, R_2) \in \mathcal{R}_O$ be achieved by a given tuple of random variables (A, B, C, X) . We show that there exists a pair of random variables (V, U) , such that $(R_{12}, R_1, R_2) \in \mathcal{C}$ and is achieved by (V, U, X) . We define (V, U) as follows. Let $\Theta \sim \text{Ber}(\lambda)$, $\lambda \in [0, 1]$, be a binary random variable independent of (A, B, C, X) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$. Define the random variable

$$\tilde{V} = \begin{cases} (A, C), & \Theta = \theta_1 \\ \emptyset, & \Theta = \theta_2 \end{cases}. \quad (110)$$

The auxiliary random variable V is defined as $V \triangleq (\Theta, \tilde{V})$. Furthermore, define

$$U = (A, B, C) \quad (111)$$

and note that (V, U) preserves the Markov structure

$$(Y_1, Y_2) - X - (U, V) \quad (112)$$

since, as stated in Lemma 17, $(Y_1, Y_2) - X - (A, B, C)$ forms a Markov chain.

First, consider the case when

$$I(A, C; Y_1) - I(A, C; Y_2) \leq 0. \quad (113)$$

By setting $\lambda = 1$ we have

$$I(V; Y_1) - I(V; Y_2) = I(A, C; Y_1) - I(A, C; Y_2) \leq 0 \stackrel{(a)}{\leq} R_{12} \quad (114)$$

where (a) follows since $R_{12} \geq 0$, which establishes (17a). (17b) holds since $H(Y_1|B, C) \leq H(Y_1)$.

We proceed with deriving (17c). The definition of (V, U) in (110)-(111) implies that

$$(A, B, C, X, Y_1, Y_2) - U - V \quad (115)$$

forms a Markov chain. Consequently, we obtain

$$I(V, U; Y_2) = I(A, B, C; Y_2) \quad (116)$$

which yields

$$I(V, U; Y_2) + R_{12} \stackrel{(a)}{=} I(A, C, B; Y_2) + R_{12} \geq I(B; Y_2|A) + R_{12} \stackrel{(b)}{\geq} R_2. \quad (117)$$

Inequality (a) in (117) follows from (111) and (115), while (b) follows from (101c). This shows that (17c) also holds.

For the sum rate, we rewrite (17d) as

$$H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) = H(Y_1|V, U) + I(V, U; Y_2) + I(V; Y_1) - I(V; Y_2) \quad (118)$$

and obtain an explicit expression for each of the information measures in the RHS of (118) in terms of (A, B, C, X) .

Based on similar arguments to those presented before, we have

$$H(Y_1|V, U) = H(Y_1|A, B, C) \quad (119)$$

while the other two information measures in (118) were previously evaluated in (114) and (116). Inserting (114), (116) and (119) into (118) results in

$$\begin{aligned} H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) &= H(Y_1|A, B, C) + I(A, B, C; Y_2) + I(A, C; Y_1) - I(A, C; Y_2) \\ &\stackrel{(a)}{\geq} H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C) \stackrel{(b)}{\geq} R_1 + R_2 \end{aligned} \quad (120)$$

where (a) follows because $\lambda = 1$ and by the mutual information chain rule, while (b) follows by (101d). This satisfies (17d).

To conclude the proof it is left to consider the case where

$$I(A, C; Y_1) - I(A, C; Y_2) > 0. \quad (121)$$

This time set

$$\lambda = \min \left\{ 1, \left(\frac{I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2)}{I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1)} \right)^+ \right\} \quad (122)$$

where $(x)^+ = \max\{0, x\}$, and consider the following.

$$I(V; Y_1) - I(V; Y_2) = \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right] \quad (123)$$

$$= \lambda \left[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1) \right] \quad (124)$$

$$\stackrel{(a)}{\leq} I(A; Y_1|C) - I(C; Y_2|A) \stackrel{(b)}{\leq} R_{12} \quad (125)$$

where (a) follows by (122) and (b) follows from (101a). Inequality (a) in (125) is proved as follows. If $\lambda = 1$ we have

$$I(A; Y_2) \geq I(C; Y_1). \quad (126)$$

Using (126), we rewrite (124) as

$$\begin{aligned} \lambda \left[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1) \right] &\stackrel{(a)}{=} I(A; Y_1|C) - I(C; Y_2|A) + I(C; Y_1) - I(A; Y_2) \\ &\stackrel{(b)}{\leq} I(A; Y_1|C) - I(C; Y_2|A) \end{aligned}$$

where (a) follows because $\lambda = 1$ and by the mutual information chain rule, while (b) follows from (126). On the

other hand, if

$$\lambda = \frac{I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2)}{I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1)} \quad (127)$$

then

$$I(A; Y_2) < I(C; Y_1) \quad (128)$$

and we rewrite (124) as

$$\begin{aligned} \lambda \left[I(A; Y_1|C) - I(A, C; Y_2) + I(C; Y_1) \right] &\stackrel{(a)}{=} I(A; Y_1|C) - I(A, C; Y_2) + I(A; Y_2) \\ &= I(A; Y_1|C) - I(C; Y_2|A). \end{aligned}$$

Here (a) follows from (127). The case $\lambda = 0$ is trivial, and we omit the derivation of (a) in (125). We conclude that (17a) is satisfied. (17b)-(17c) follow by the same arguments presented above, while for (17d) we have

$$\begin{aligned} H(Y_1|V, U) + I(U; Y_2|V) + I(V; Y_1) &\stackrel{(a)}{=} H(Y_1|A, B, C) + I(A, B, C; Y_2) + \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right] \\ &\stackrel{(b)}{\geq} H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C) \stackrel{(c)}{\geq} R_1 + R_2 \end{aligned} \quad (129)$$

where (a) follows by (115) and (118), (b) follows from the choice of λ in (122), while (c) follows by (101d).

The derivation of (b) in (129) relies on evaluating the terms of interest for the three possible values of λ . First, by (121), $\lambda = 0$ if and only if

$$I(A; Y_1|C) \leq I(C; Y_2|A) \quad (130)$$

which implies

$$\begin{aligned} H(Y_1|A, B, C) + I(A, B, C; Y_2) + \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right] \\ &= H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A, C; Y_2) \\ &\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(C; Y_2|A) \\ &\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_2|C). \end{aligned} \quad (131)$$

If $\lambda = 1$, by using the mutual information chain rule we have

$$\begin{aligned} H(Y_1|A, B, C) + I(A, B, C; Y_2) + \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right] \\ &= H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A, C; Y_1) - I(A, C; Y_2) \\ &\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C). \end{aligned} \quad (132)$$

Finally, if λ satisfies (127), we obtain

$$H(Y_1|A, B, C) + I(A, B, C; Y_2) + \lambda \left[I(A, C; Y_1) - I(A, C; Y_2) \right]$$

$$\begin{aligned}
&= H(Y_1|A, B, C) + I(A; Y_2) + I(B; Y_2|A, C) + I(A; Y_1|C) \\
&\geq H(Y_1|A, B, C) + I(B; Y_2|A, C) + I(A; Y_1|C).
\end{aligned} \tag{133}$$

We find that (17d) is satisfied as well. Concluding, (17) holds for the choice of (V, U) and λ stated in (110)-(111) and (122), respectively. This implies that $\mathcal{R}_O \subseteq \mathcal{C}$. ■

Lemma 18 completes the converse and characterizes the region in (13) as the capacity region of the SD-BC with cooperation.

Remark 19 *The definition of V in (110) is probabilistic and depends on the joint distribution of (A, B, C, X) that is induced by the codebook. More specifically, the choice of λ varies with each joint distribution and maintains that the inclusion in (109) is valid.*

APPENDIX F

DERIVATION OF THE REGION IN (13) FROM (25)

Denote the region in (25) by \mathcal{R} . Note that \mathcal{C}_{BC} is achievable from \mathcal{R} by taking X_1 to be independent of (V, U, X) and applying a coding scheme where the transmission rate via the relay channel is R_{12} . This implies that $\mathcal{C}_{BC} \subseteq \mathcal{R}$.

To show that $\mathcal{R} \subseteq \mathcal{C}_{BC}$ recall that the proof of Theorem 8 in [35] relies on Theorem 4 in that same work, which characterized an upper bound on the capacity region of a general RBC. In the proof of Theorem 4 (see [35, Appendix II]), the auxiliary random variables V_i and U_i are defined as

$$V_i \triangleq (M_0, Y_1^{i-1}, Y_{2,i+1}^n) \quad ; \quad U_i \triangleq (M_2, Y_1^{i-1}, Y_{2,i+1}^n). \tag{134}$$

M_0 is a common message that was also considered in [35]. Since $X_{1,i}$ is a function of Y_1^{i-1} , it is also a function of V_i (and/or U_i) for every $i \in [1 : n]$. In particular, this implies that X_1 is a function of V . The information measures defining \mathcal{R} are then upper bounded as follows. For R_1 we have

$$R_1 \leq H(Y_1|X_1) \leq H(Y_1). \tag{135}$$

For the R_2 consider

$$\begin{aligned}
R_2 &\leq I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21}) \\
&\stackrel{(a)}{\leq} I(V, U; Y_{21}) + H(Y_{22}) \\
&\stackrel{(b)}{\leq} I(V, U; Y_{21}) + R_{12}
\end{aligned} \tag{136}$$

where (a) follows because X_1 is a function of V and since conditioning cannot increase entropy, while (b) follows because the relay channel is deterministic with capacity R_{12} .

For the first bound on the sum of rates, we have

$$\begin{aligned} R_1 + R_2 &\leq H(Y_1|V, U, X_1) + I(U; Y_{21}|V, X_1) + I(V; Y_1|X_1) \\ &\stackrel{(a)}{\leq} H(Y_1|V, U) + I(U; Y_{21}|V) + I(V; Y_1). \end{aligned} \quad (137)$$

Here (a) is justified similarly to step (a) in (136).

Finally, the second bound on $R_1 + R_2$ is upper bounded as

$$\begin{aligned} R_1 + R_2 &\leq H(Y_1|V, U, X_1) + I(V, U, X_1; Y_{21}) + H(Y_{22}|Y_{21}) \\ &\stackrel{(a)}{\leq} H(Y_1|V, U) + I(V, U; Y_{21}) + H(Y_{22}) \\ &\stackrel{(b)}{\leq} H(Y_1|V, U) + I(V, U; Y_{21}) + R_{12}. \end{aligned} \quad (138)$$

Again, (a) and (b) follows by the same arguments as (a) and (b) in (136).

To complete the proof, it remains to be shown that taking the union only over PMFs in which X_1 is independent of (V, U, X) , exhausts the entire region. This follows since the rate bounds in (135)-(138) do not involve X_1 nor Y_{22} . Relabeling Y_{21} as Y_2 yields $\mathcal{R} \subseteq \mathcal{C}_{BC}$ and completes the proof.

APPENDIX G DERIVATION OF THE REGION IN (27)

We prove that the admissible region for the SW problem with one-sided encoder cooperation is (27), and that (27) is obtained from \mathcal{R}_{WAK} stated in Theorem 6. To achieve (27) from \mathcal{R}_{WAK} , we set $U = X_2$ and evaluate the rate bounds in (10) to get

$$\begin{aligned} R_{12} &\geq I(V; X_1|X_2) \\ R_1 &\geq H(X_1|X_2) - I(V; X_1|X_2) \\ R_2 &\geq H(X_2|X_1) \\ R_1 + R_2 &\geq H(X_1, X_2). \end{aligned} \quad (139)$$

The structure of the region in (27) implies $R_{12} \leq H(X_1|X_2)$. Thus, it suffices to show that for every $0 \leq R_{12} \leq H(X_1|X_2)$ there is a random variable V that admits the Markov property $V - X_1 - X_2$ such that $I(V; X_1|X_2) = R_{12}$. Since $R_{12} \leq H(X_1|X_2)$, there is a real number $\gamma \geq 0$ such that

$$R_{12} = H(X_1|X_2) - \gamma. \quad (140)$$

Set the auxiliary random variable $V \triangleq (\Theta, \tilde{V})$, where $\Theta \sim \text{Ber}(\lambda)$, $\lambda \in [0, 1]$, is a binary random variable

independent of (X_1, X_2) that takes values in $\mathcal{O} = \{\theta_1, \theta_2\}$, and

$$\tilde{V} = \begin{cases} X_1, & \Theta = \theta_1 \\ \emptyset, & \Theta = \theta_2 \end{cases}. \quad (141)$$

Taking

$$\lambda = \frac{H(X_1|X_2) - \gamma}{H(X_1|X_2)}, \quad (142)$$

results in

$$\begin{aligned} I(V; X_1|X_2) &= \lambda I(X_1; X_1|X_2) + \bar{\lambda} I(\emptyset; X_1|X_2) \\ &= \lambda H(X_1|X_2) + \bar{\lambda} \cdot 0 \\ &= H(X_1|X_2) - \gamma = R_{12}. \end{aligned}$$

This implies the achievability of (27).

The converse follows by the generalized Cut-Set bound [58, Theorem 1] and characterizes (27) as the admissible rate region for the SW problem with one-sided encoder cooperation.

APPENDIX H

PROOF OF THE MARKOV RELATION IN (85)

We present two proofs for the Markov relation in (85), each based on a different graphical method. The first uses the sufficient condition via undirected graphs that was introduced in [59]. The second approach relies on the notion of d-separation in functional dependence graphs (FDGs), for which we use the formulation from [60].

By the definitions of the auxiliaries V and U , it suffices to show that

$$(M_1, M_2, M_{12}, Y_1^{q-1}, Y_{2,q+1}^n, Y_{1,q}) - X_q - Y_{2,q} \quad (143)$$

is a Markov chain for every $q \in [1 : n]$. In fact, we prove the stronger Markov relation

$$(M_1, M_2, Y_1^n, Y_{2,q+1}^n) - X_q - Y_{2,q} \quad (144)$$

from which (143) follows because M_{12} is a function of Y_1^n . Since the channel is SD, memoryless and without feedback, for every $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, $(x^n, y_1^n, y_2^n) \in \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ and $q \in [1 : n]$, we have

$$\begin{aligned} P(m_1, m_2, x^n, y_1^n, y_2^n) &= P(m_1)P(m_2)P(x^n|m_1, m_2)P(y_1^{q-1}|x^{q-1})P(y_2^{q-1}|x^{q-1}) \\ &\quad \times P(y_{1,q}|x_q)P(y_{2,q}|x_q)P(y_{1,q+1}^n|x_{q+1}^n)P(y_{2,q+1}^n|x_{q+1}^n) \end{aligned} \quad (145)$$

Given (145), the Markov relation in (144) follows by using either of two subsequently explained methods.

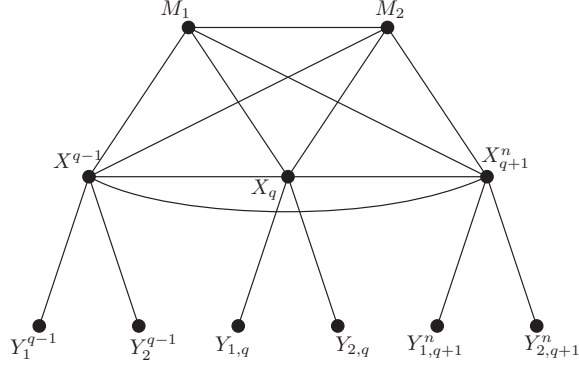


Fig. 6: Undirected graph corresponding to the PMF in (145): the relation (144) holds because all paths from $Y_{2,q}$ to $(M_1, M_2, Y_1^n, Y_{2,q+1}^n)$ pass through X_q .

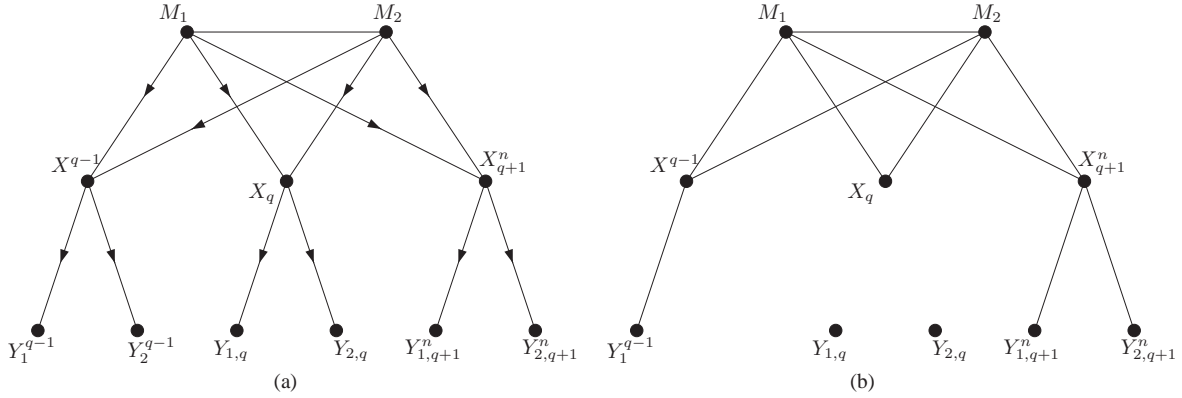


Fig. 7: (a) The FDG that stems from (145): (144) follows since $\mathcal{C} = \{X_q\}$ d-separates $\mathcal{A} = \{Y_{2,q}\}$ from $\mathcal{B} = \{M_1, M_2, Y_1^n, Y_{2,q+1}^n\}$. (b) The undirected graph obtained from the FDG after the manipulations described in Definition 10.

A. Via Undirected Graph

Fig. 6 shows the undirected graph that stems from the PMF in (145) with respect to the principles described in [59]. Namely, the nodes of the graph correspond to the random variables in (145). All the nodes that are associated with random variables that appear together in any of the terms in the factorization of (145) are connected by edges. For instance, the term $P(x^n|m_1, m_2)$ induces edges that connect the nodes of M_1 , M_2 , X^{q-1} , X_q and X_{q+1}^n with one another. The Markov chain in (144) follows from Fig. 6, since all paths from $Y_{2,q}$ to $(M_1, M_2, Y_1^n, Y_{2,q+1}^n)$ pass through X_q .

B. Via Functional Dependence Graph and d-Separation

Fig. 7(a) shows the FDG induced from (145). The structure of FDGs allows one to establish the conditional statistical independence of sets of random variables using the notion of d-separation.

Definition 10 (d-separation [60]) *Let \mathcal{A} , \mathcal{B} and \mathcal{C} be disjoint subsets of the vertices of an FDG \mathcal{G} . \mathcal{C} is said to d-separate \mathcal{A} from \mathcal{B} if there is no path between a vertex in \mathcal{A} and a vertex in \mathcal{B} after the following manipulations of the graph have been performed.*

- 1) *Consider the subgraph $\mathcal{G}_{\mathcal{ABC}}$ of \mathcal{G} consisting of the vertices in \mathcal{A} , \mathcal{B} and \mathcal{C} , as well as the edges and vertices encountered when moving backward one or more edges starting from any of the vertices in \mathcal{A} , \mathcal{B} or \mathcal{C} .*
- 2) *In $\mathcal{G}_{\mathcal{ABC}}$, delete all edges coming out of the vertices in \mathcal{C} . Call the resulting graph $\mathcal{G}_{\mathcal{AB}|\mathcal{C}}$.*
- 3) *Remove the arrows on the remaining edges of $\mathcal{G}_{\mathcal{AB}|\mathcal{C}}$ to obtain an undirected graph.*

The Markov relation in (145) follows by setting $\mathcal{A} = \{Y_{2,q}\}$, $\mathcal{B} = \{M_1, M_2, Y_1^n, Y_{2,q+1}^n\}$ and $\mathcal{C} = \{X_q\}$, and noting that \mathcal{C} d-separates \mathcal{A} from \mathcal{B} [60]. To see this, in Fig. 7(b) we illustrate the undirected graph obtained from the FDG in Fig. 7(a) by applying the manipulations described in Definition 10 with respect to the specified choices of \mathcal{A} , \mathcal{B} and \mathcal{C} .

Neither of the methods is a special case of the other. While the first method (via undirected graphs) involves graphs with more edges, the derivation of the Markov relations using such graphs is more direct. The second method (via FDGs and d-separation) requires manipulating the originally constructed FDG. However, the FDG is typically simpler than its undirected counterpart.

REFERENCES

- [1] F. M. J. Willems. The discrete memoryless multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory*, 29(6):441–445, May 1983.
- [2] T. Berger. Multiterminal source coding. In G. Longo, editor, *Information Theory Approach to Communications*, volume 229, pages 171–231. CISM Course and Lecture, 1978.
- [3] S. Y. Tung. *Multiterminal source coding*. PhD thesis, Cornell University, New York, United States, May 1978.
- [4] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4):471–480, Jul. 1973.
- [5] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 22(1):1–10, Jan. 1976.
- [6] A. B. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic gaussian two-encoder source-coding problem. *IEEE Trans. Inf. Theory*, 54(5):1938–1961, May 2008.
- [7] A. D. Wyner. On source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 21(3):294–300, May 1975.
- [8] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. Inf. Theory*, 21(6):629–637, 1975.
- [9] A. D. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2):163–179, Mar. 1975.
- [10] G. Kramer and S. A. Savari. Data compression with commuting density operators. In *IEEE Int. Symp. Inf. Theory*, page 115, Lausanne, Switzerland, Jun.-Jul. 2002.
- [11] G. Kramer and S. A. Savari. Communicating probability distributions. *IEEE Trans. Inf. Theory*, 53(2):518–525, Feb. 2007.
- [12] V. Anantharam and V. Borkar. Common randomness and distributed control; A counterexample. *Syst. Control Lett.*, 56:568–572, 2007.
- [13] P. W. Cuff, H. H. Permuter, and T. M. Cover. Coordination capacity. *IEEE Trans. Inf. Theory*, 56(9):4181–4206, Sep. 2010.
- [14] A. Berezni, M. Bahrani, M. Mirmohseni, and M. R. Aref. Empirical coordination in a triangular multiterminal network. In *In. Proc. Symp. Inf. theory*, Istanbul, Turkey, Jul. 2013.
- [15] H. Asnani, H. H. Permuter, and T. Weissman. Successive refinement with decoder cooperation and its channel coding duals. *IEEE Trans. Inf. Theory*, 59(9):5511–5533, Sep 2013.
- [16] T. M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Trans. Inf. Theory*, 22(2):226–228, Mar. 1975.
- [17] T. M. Cover. Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1):2–14, Jan. 1972.
- [18] T. Berger and R.W. Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Trans. Inf. Theory*, 35(2):228–236, Mar. 1989.
- [19] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, pages 142–163, 1959.
- [20] T. M. Cover and M. Chiang. Unified duality between channel capacity and rate distortion with state information. *IEEE Trans. Inf. Theory*, 48(6):1629–1638, Jun. 2002.
- [21] S. S. Pradhan, J. Chou, and K. Ramchandran. Duality between source coding and channel coding and its extension to the side information case. *IEEE Trans. Inf. Theory*, 49(5):1181–1203, May 2003.
- [22] A. Gupta and S. Verdú. Operational duality between lossy compression and channel coding. *IEEE Trans. Inf. Theory*, 57(6):3171–3179, Jun 2011.
- [23] W. Yu. Duality and the value of cooperation in distributed source and channel coding problems. In *Proc. 41st Annu. Allerton Conf. Commun., Control and Comput.*, Monticell, Illinois, United States, 2003.
- [24] H. Wang and P. Viswanath. Fixed binning schemes: An operational duality between channel and source coding problems with side information. In *in Proc. Int. Symp. Inf. Theory*, Chicago, Illinois, United States, Jun. 2004.
- [25] L. Dikstein, H. H. Permuter, and S. Shamai (Shitz). MAC with action-dependent state information at one encoder. *Submitted for publication to IEEE Trans. Inf. Theory*, 2012. <http://arxiv.org/abs/1212.4626>.
- [26] S. I. Bross, A. Lapidoth, and M. A. Wigger. The Gaussian MAC with conferencing encoders. In *Proc. Int. Symp. Inf. Theory*, pages 2702–2706, Toronto, Canada, Jul. 2008.
- [27] O. Simeone, D. Gündüz, H. V. Poor, A. J. Goldsmith, and S. Shamai (Shitz). Compound multiple-access channels with partial cooperation. *IEEE Trans. Inf. Theory*, 55(6):2425–2441, Jun 2009.

- [28] M. Wiese, H. Boche, I. Bjelakovic, and V. Jungnickel. The compound multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory*, 57(5):3045–3066, May 2011.
- [29] R. Dabora and S. D. Servetto. Broadcast channels with cooperating decoders. *IEEE Trans. Inf. Theory*, 52:5438–5454, 2006.
- [30] Y. Liang and V. V. Veeravalli. Cooperative relay broadcast channels. *IEEE Trans. Inf. Theory*, 51(3):900–928, Mar 2007.
- [31] S. I. Gelfand and M. S. Pinsker. Capacity of a broadcast channel with one deterministic component. *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, 16(1):17–25, Jan-Mar 1980.
- [32] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, 25(3):306–311, May 1979.
- [33] A. Lapidoth and L. Wang. The state-dependent semideterministic broadcast channel. *IEEE Trans. Inf. Theory*, 59(4):2242–2251, Apr. 2013.
- [34] G. Kramer. Teaching IT: An identity for the Gelfand-Pinsker converse. *IEEE Inf. Theory Society Newsletter*, 61(4):4–6, Dec. 2011.
- [35] Y. Liang and G. Kramer. Rate regions for relay broadcast channels. *IEEE Trans. Inf. Theory*, 53(10):3517–3535, Oct 2007.
- [36] R. Kötter, M. Effros, and M. Médard. Theory of network equivalence - Part I: Point-to-point channels. *IEEE Trans. Inf. Theory*, 57:972–995, Feb. 2011.
- [37] H. H. Permuter, Z. Goldfeld and G. Kramer. The Ahlswede-Körner coordination problem with one-sided encoder cooperation. In *Proc. Int. Symp. Inf. Theory (ISIT-2014)*, Honolulu, Hawaii, US, Jun.-Jul. 2014.
- [38] Z. Goldfeld, H. H. Permuter, and G. Kramer. Duality of a source coding problem and the semi-deterministic broadcast channel with rate-limited cooperation. *Submitted for publication to IEEE Trans. Inf. Theory, passed 1st round, currently undergoing review prior to publication*, 2014. Available on ArXiv at <http://arxiv.org/abs/1405.7812>.
- [39] R. Dabora and S. Servetto. On the role of estimate-and-forward with time sharing in cooperative communication. *IEEE Trans. Inf. Theory*, 54(10):4409–4431, Oct. 2008.
- [40] T. A. Courtade and T. Weissman. Multiterminal source coding under logarithmic loss. *IEEE Trans. Inf. Theory*, 60(1):740–761, Jan. 2014.
- [41] L. Dikstein, H. H. Permuter, and Y. Steinberg. On state dependent broadcast channels with cooperation. *Submitted for publication to IEEE Trans. Inf. Theory*, 2014.
- [42] J. L. Massey. *Applied Digital Information Theory*. ETH Zurich, Zurich, Switzerland, 1980-1998.
- [43] A. Orlitsky and J. Roche. Coding for computing. *IEEE Trans. Inf. Theory*, 47(3):903–917, Mar 2001.
- [44] Csiszár and Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge Univ. Press, 2nd edition, 2011.
- [45] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New-York, 2nd edition, 2006.
- [46] G. Dueck and J. Körner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Trans. Inf. Theory*, 25(1):82–85, Jan 1979.
- [47] J. Pokorny and H.-M. Wallmeier. Random coding bound and codes produced by permutations for the multiple-access channel. *IEEE Trans. Inf. Theory*, 31(6):741–750, Nov 1985.
- [48] Y.-S. Liu and B. L. Hughes. A new universal random coding bound for the multiple access channel. *IEEE Trans. Inf. Theory*, 42(2):376–386, Mar 1996.
- [49] C. Chang. Interference channel capacity region for randomized fixed-composition codes. In *47th Annual Allerton Conference on Communication, Control, and Computing*, pages 280–287, Allerton Retreat Center, Monticello, Illinois, Sep 2009.
- [50] Z. Goldfeld, H. H. Permuter, and G. Kramer. Semi-deterministic broadcast channels with cooperation. In *Proc. 28-th Convention of Electrical and Electronics Engineers (IEEEI-2014)*, Eilat, Israel, Dec. 2014.
- [51] J. Körner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory*, 23(1):60–64, Jan. 1977.
- [52] A. El Gamal. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory*, 25(2):166–169, Mar. 1979.
- [53] T. M. Cover and Y.-H. Kim. Capacity of a class of deterministic relay channels. *IEEE Trans. Inf. Theory*, 54(3):1328–1329, Feb. 2008.
- [54] S. C. Draper, B. J. Frey, and F. R. Kschischang. On interacting encoders and decoders in multiuser settings. In *Proc. Int. Symp. Inf. Theory (ISIT-2004)*, Chicago, Illinois, US, Jun.-Jul. 2004.
- [55] T.D. Nguyen and S. Lasaulce. Capacity region of the deterministic broadcast channel with cooperative decoders. Technical report, École supérieure d’électricité, Rennes, Metz, France, October 2005.
- [56] A. H. Kaspi. Two-way source coding with a fidelity criterion,. *IEEE Trans. Inf. Theory*, 31(6):735–740, November 1985.
- [57] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

- [58] A. Cohen, S. Avestimehr, and M. Effros. On networks with side information. In *Proc. Int. Symp. Inf. Theory (ISIT-2009)*, Seoul, Korea, Jun.-Jul. 2009.
- [59] H. H. Permuter, Y. Steinberg, and T. Weissman. Two-way source coding with a helper. *IEEE Trans. Inf. Theory*, 56(6):2905–2919, Jun. 2010.
- [60] G. Kramer. Capacity results for the discrete memoryless networks. *IEEE. Trans. Inf. Theory*, 49(1):4–21, Jan. 2003.